RECEPTORS AND MEMBRANE-ASSOCIATED PROTEINS

TECHNICAL FIELD

The invention relates to novel nucleic acids, receptors and membrane-associated proteins encoded by these nucleic acids, and to the use of these nucleic acids and proteins in the diagnosis, treatment, and prevention of cell proliferative, autoimmune/inflammatory, neurological, metabolic, developmental, and endocrine disorders. The invention also relates to the assessment of the effects of exogenous compounds on the expression of nucleic acids and receptors and membrane-associated proteins.

BACKGROUND OF THE INVENTION

Signal transduction is the general process by which cells respond to extracellular signals. Signal transduction across the plasma membrane begins with the binding of a signal molecule, e.g., a hormone, neurotransmitter, or growth factor, to a cell membrane receptor. The receptor, thus activated, triggers an intracellular biochemical cascade that ends with the activation of an intracellular target molecule, such as a transcription factor. This process of signal transduction regulates all types of cell functions including cell proliferation, differentiation, and gene transcription.

Biological membranes surround organelles, vesicles, and the cell itself. Membranes are highly selective permeability barriers made up of lipid bilayer sheets composed of phosphoglycerides, fatty acids, cholesterol, phospholipids, glycolipids, proteoglycans, and proteins. Membranes contain ion pumps, ion channels, and specific receptors for external stimuli which transmit biochemical signals across the membranes. These membranes also contain second messenger proteins which interact with these pumps, channels, and receptors to amplify and regulate transmission of these signals.

Plasma Membrane Proteins

Plasma membrane proteins (MPs) are divided into two groups based upon methods of protein extraction from the membrane. Extrinsic or peripheral membrane proteins can be released using extremes of ionic strength or pH, urea, or other disruptors of protein interactions. Intrinsic or integral membrane proteins are released only when the lipid bilayer of the membrane is dissolved by detergent.

The majority of known integral membrane proteins are transmembrane proteins (TM) which are characterized by an extracellular, a transmembrane, and an intracellular domain. TM domains are typically comprised of 15 to 25 hydrophobic amino acids which are predicted to adopt an α-helical conformation. TM proteins are classified as bitopic (Types I and II) and polytopic (Types III and IV) (Singer, S.J. (1990) Annu. Rev. Cell Biol. 6:247-296). Bitopic proteins span the membrane once

while polytopic proteins contain multiple membrane-spanning segments. TM proteins carry out a variety of important cellular functions, including acting as cell-surface receptor proteins involved in signal transduction. These functions are represented by growth and differentiation factor receptors, and receptor-interacting proteins such as *Drosophila* pecanex and frizzled proteins, LIV-1 protein, NF2 protein, and GNS1/SUR4 eukaryotic integral membrane proteins. TM proteins also act as transporters of ions or metabolites, such as gap junction channels (connexins), and ion channels, and as cell anchoring proteins, such as lectins, integrins, and fibronectins. TM proteins may be vesicle organelle-forming molecules, such as caveolins, or cell recognition molecules, such as cluster of differentiation (CD) antigens, glycoproteins, and mucins.

Many MPs contain amino acid sequence motifs that serve to localize proteins to specific subcellular sites. Examples of these motifs include PDZ domains, KDEL, RGD, NGR, and GSL sequence motifs, von Willebrand factor A (vWFA) domains, and EGF-like domains. RGD, NGR, and GSL motif-containing peptides have been used as drug delivery agents in targeted cancer treatment of tumor vasculature (Arap, W. et al. (1998) Science, 279:377-380). Furthermore, MPs may also contain amino acid sequence motifs that serve to interact with extracellular or intracellular molécules, such as carbohydrate recognition domains (CRD).

Chemical modification of amino acid residue side chains alters the manner in which MPs interact with other molecules, for example, phospholipid membranes. Examples of such chemical modifications to amino acid residue side chains are covalent bond formation with glycosaminoglycans, oligosaccharides, phospholipids, acetyl and palmitoyl moieties, ADP-ribose, phosphate, and sulphate groups.

RNA encoding membrane proteins may have alternative splice sites which give rise to proteins encoded by the same gene but with different messenger RNA and amino acid sequences. Splice variant membrane proteins may interact with other ligand and protein isoforms.

Receptors

The term receptor describes proteins that specifically recognize other molecules. The category is broad and includes proteins with a variety of functions. The bulk of receptors are cell surface proteins which bind extracellular ligands and produce cellular responses in the areas of growth, differentiation, endocytosis, and immune response. Other receptors facilitate the selective transport of proteins out of the endoplasmic reticulum and localize enzymes to particular locations in the cell. The term may also be applied to proteins which act as receptors for ligands with known or unknown chemical composition and which interact with other cellular components. For example, the steroid hormone receptors bind to and regulate transcription of DNA.

Cell surface receptors are typically integral plasma membrane proteins. These receptors

recognize hormones such as catecholamines; peptide hormones; growth and differentiation factors; small peptide factors such as thyrotropin-releasing hormone; galanin, somatostatin, and tachykinins; and circulatory system-borne signaling molecules. Cell surface receptors on immune system cells recognize antigens, antibodies, and major histocompatibility complex (MHC)-bound peptides. Other cell surface receptors bind ligands to be internalized by the cell. This receptor-mediated endocytosis functions in the uptake of low density lipoproteins (LDL), transferrin, glucose- or mannose-terminal glycoproteins, galactose-terminal glycoproteins, immunoglobulins, phosphovitellogenins, fibrin, proteinase-inhibitor complexes, plasminogen activators, and thrombospondin (Lodish, H. et al. (1995) Molecular Cell Biology, Scientific American Books, New York NY, p. 723; Mikhailenko, I. et al. (1997) J. Biol. Chem. 272:6784-6791).

Receptor Protein Kinases

Many growth factor receptors, including receptors for epidermal growth factor, platelet-derived growth factor, fibroblast growth factor, as well as the growth modulator α-thrombin, contain intrinsic protein kinase activities. When growth factor binds to the receptor, it triggers the autophosphorylation of a serine, threonine, or tyrosine residue on the receptor. These phosphorylated sites are recognition sites for the binding of other cytoplasmic signaling proteins. These proteins participate in signaling pathways that eventually link the initial receptor activation at the cell surface to the activation of a specific intracellular target molecule. In the case of tyrosine residue autophosphorylation, these signaling proteins contain a common domain referred to as a Src homology (SH) domain. SH2 domains and SH3 domains are found in phospholipase C-γ, PI-3-K p85 regulatory subunit, Ras-GTPase activating protein, and $pp60^{c\text{-}src}$ (Lowenstein, E.J. et al. (1992) Cell 70:431-442). The cytokine family of receptors share a different common binding domain and include transmembrane receptors for growth hormone (GH), interleukins, erythropoietin, and prolactin.

Other receptors and second messenger-binding proteins have intrinsic serine/threonine protein kinase activity. These include activin/TGF-β/BMP-superfamily receptors, calcium- and diacylglycerol-activated/phospholipid-dependant protein kinase (PK-C), and RNA-dependant protein kinase (PK-R). In addition, other serine/threonine protein kinases, including nematode Twitchin, have fibronectin-like, immunoglobulin C2-like domains.

G-protein coupled receptors

The G-protein coupled receptors (GPCRs), encoded by one of the largest families of genes yet identified, play a central role in the transduction of extracellular signals across the plasma membrane. GPCRs have a proven history of being successful therapeutic targets.

GPCRs are integral membrane proteins characterized by the presence of seven hydrophobic transmembrane domains which together form a bundle of antiparallel alpha (α) helices. GPCRs range in size from under 400 to over 1000 amino acids (Strosberg, A.D. (1991) Eur. J. Biochem. 196:1-10;

Coughlin, S.R. (1994) Curr. Opin. Cell Biol. 6:191-197). The amino-terminus of a GPCR is extracellular, is of variable length, and is often glycosylated. The carboxy-terminus is cytoplasmic and generally phosphorylated. Extracellular loops alternate with intracellular loops and link the transmembrane domains. Cysteine disulfide bridges linking the second and third extracellular loops may interact with agonists and antagonists. The most conserved domains of GPCRs are the transmembrane domains and the first two cytoplasmic loops. The transmembrane domains account, in part, for structural and functional features of the receptor. In most cases, the bundle of α helices forms a ligand-binding pocket. The extracellular N-terminal segment, or one or more of the three extracellular loops, may also participate in ligand binding. Ligand binding activates the receptor by inducing a conformational change in intracellular portions of the receptor. In turn, the large, third intracellular loop of the activated receptor interacts with a heterotrimeric guanine nucleotide binding (G) protein complex which mediates further intracellular signaling activities, including the activation of second messengers such as cyclic AMP (cAMP), phospholipase C, and inositol triphosphate, and the interaction of the activated GPCR with ion channel proteins. (See, e.g., Watson, S. and S. Arkinstall (1994) The G-protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 2-6; Bolander, F.F. (1994) Molecular Endocrinology, Academic Press, San Diego CA, pp. 162-176; Baldwin, J.M. (1994) Curr. Opin. Cell Biol. 6:180-190.)

GPCRs include receptors for sensory signal mediators (e.g., light and olfactory stimulatory molecules); adenosine, γ-aminobutyric acid (GABA), hepatocyte growth factor, melanocortins, neuropeptide Y, opioid peptides, opsins, somatostatin, tachykinins, vasoactive intestinal polypeptide family, and vasopressin; biogenic amines (e.g., dopamine, epinephrine and norepinephrine, histamine, glutamate (metabotropic effect), acetylcholine (muscarinic effect), and serotonin); chemokines; lipid mediators of inflammation (e.g., prostaglandins and prostanoids, platelet activating factor, and leukotrienes); and peptide hormones (e.g., bombesin, bradykinin, calcitonin, C5a anaphylatoxin, endothelin, follicle-stimulating hormone (FSH), gonadotropic-releasing hormone (GnRH), neurokinin, and thyrotropin-releasing hormone (TRH), and oxytocin). GPCRs which act as receptors for stimuli that have yet to be identified are known as orphan receptors.

GPCR mutations, which may cause loss of function or constitutive activation, have been associated with numerous human diseases (Coughlin, *supra*). For instance, retinitis pigmentosa may arise from mutations in the rhodopsin gene. Furthermore, somatic activating mutations in the thyrotropin receptor have been reported to cause hyperfunctioning thyroid adenomas, suggesting that certain GPCRs susceptible to constitutive activation may behave as protooncogenes (Parma, J. et al. (1993) Nature 365:649-651). GPCR receptors for the following ligands also contain mutations associated with human disease: luteinizing hormone (precocious puberty); vasopressin V_2 (X-linked nephrogenic diabetes); glucagon (diabetes and hypertension); calcium (hyperparathyroidism,

hypocalcuria, hypercalcemia); parathyroid hormone (short limbed dwarfism); β_3-adrenoceptor (obesity, non-insulin-dependent diabetes mellitus); growth hormone releasing hormone (dwarfism); and adrenocorticotropin (glucocorticoid deficiency) (Wilson, S. et al. (1998) Br. J. Pharmocol. 125:1387-1392; Stadel, J.M. et al. (1997) Trends Pharmacol. Sci. 18:430-437). GPCRs are also involved in depression, schizophrenia, sleeplessness, hypertension, anxiety, stress, renal failure, and several cardiovascular disorders (Horn, F. and G. Vriend (1998) J. Mol. Med. 76:464-468).

In addition, within the past 20 years several hundred new drugs have been recognized that are directed towards activating or inhibiting GPCRs. The therapeutic targets of these drugs span a wide range of diseases and disorders, including cardiovascular, gastrointestinal, and central nervous system disorders as well as cancer, osteoporosis and endometriosis (Wilson et al., *supra*; Stadel et al., *supra*). For example, the dopamine agonist L-dopa is used to treat Parkinson's disease, while a dopamine antagonist is used to treat schizophrenia and the early stages of Huntington's disease. Agonists and antagonists of adrenoceptors have been used for the treatment of asthma, high blood pressure, other cardiovascular disorders, and anxiety; muscarinic agonists are used in the treatment of glaucoma and tachycardia; serotonin 5HT1D antagonists are used against migraine; and histamine H1 antagonists are used against allergic and anaphylactic reactions, hay fever, itching, and motion sickness (Horn et al., *supra*).

Nuclear Receptors

Nuclear receptors bind small molecules such as hormones or second messengers, leading to increased receptor-binding affinity to specific chromosomal DNA elements. In addition the affinity for other nuclear proteins may also be altered. Such binding and protein-protein interactions may regulate and modulate gene expression. Examples of such receptors include the steroid hormone receptors family, the retinoic acid receptors family, and the thyroid hormone receptors family.

Ligand-Gated Receptor Ion Channels

Ligand-gated receptor ion channels fall into two categories. The first category, extracellular ligand-gated receptor ion channels (ELGs), rapidly transduce neurotransmitter-binding events into electrical signals, such as fast synaptic neurotransmission. ELG function is regulated by post-translational modification. The second category, intracellular ligand-gated receptor ion channels (ILGs), are activated by many intracellular second messengers and do not require post-translational modification(s) to effect a channel-opening response.

ELGs depolarize excitable cells to the threshold of action potential generation. In non-excitable cells, ELGs permit a limited calcium ion-influx during the presence of agonist. ELGs include channels directly gated by neurotransmitters such as acetylcholine, L-glutamate, glycine, ATP, serotonin, GABA, and histamine. ELG genes encode proteins having strong structural and functional similarities. ILGs are encoded by distinct and unrelated gene families and include

receptors for cAMP, cGMP, calcium ions, ATP, and metabolites of arachidonic acid.

Macrophage Scavenger Receptors

Macrophage scavenger receptors with broad ligand specificity may participate in the binding of low density lipoproteins (LDL) and foreign antigens. Scavenger receptors types I and II are trimeric membrane proteins with each subunit containing a small N-terminal intracellular domain, a transmembrane domain, a large extracellular domain, and a C-terminal cysteine-rich domain. The extracellular domain contains a short spacer domain, an α-helical coiled-coil domain, and a triple helical collagenous domain. These receptors have been shown to bind a spectrum of ligands, including chemically modified lipoproteins and albumin, polyribonucleotides, polysaccharides, phospholipids, and asbestos (Matsumoto, A. et al. (1990) Proc. Natl. Acad. Sci. USA 87:9133-9137; Elomaa, O. et al. (1995) Cell 80:603-609). The scavenger receptors are thought to play a key role in atherogenesis by mediating uptake of modified LDL in arterial walls, and in host defense by binding bacterial endotoxins, bacteria, and protozoa.

T-Cell Receptors

T cells play a dual role in the immune system as effectors and regulators, coupling antigen recognition with the transmission of signals that induce cell death in infected cells and stimulate proliferation of other immune cells. Although a population of T cells can recognize a wide range of different antigens, an individual T cell can only recognize a single antigen and only when it is presented to the T cell receptor (TCR) as a peptide complexed with a major histocompatibility molecule (MHC) on the surface of an antigen presenting cell. The TCR on most T cells consists of immunoglobulin-like integral membrane glycoproteins containing two polypeptide subunits, α and β, of similar molecular weight. Both TCR subunits have an extracellular domain containing both variable and constant regions, a transmembrane domain that traverses the membrane once, and a short intracellular domain (Saito, H. et al. (1984) Nature 309:757-762). The genes for the TCR subunits are constructed through somatic rearrangement of different gene segments. Interaction of antigen in the proper MHC context with the TCR initiates signaling cascades that induce the proliferation, maturation, and function of cellular components of the immune system (Weiss, A. (1991) Annu. Rev. Genet. 25:487-510). Rearrangements in TCR genes and alterations in TCR expression have been noted in lymphomas, leukemias, autoimmune disorders, and immunodeficiency disorders (Aisenberg, A.C. et al. (1985) N. Engl. J. Med. 313:529-533; Weiss, *supra*).

Netrin Receptors:

The netrins are a family of molecules that function as diffusible attractants and repellants to guide migrating cells and axons to their targets within the developing nervous system. The netrin receptors include the *C. elegans* protein UNC-5, as well as homologues recently identified in vertebrates (Leonardo, E.D. et al. (1997) Nature 386:833-838). These receptors are members of the

immunoglobulin superfamily, and also contain a characteristic domain called the ZU5 domain. Mutations in the mouse member of the netrin receptor family, Rcm (rostral cerebellar malformation) result in cerebellar and midbrain defects as an apparent result of abnormal neuronal migration (Ackerman, S.L. et al. (1997) Nature 386:838-842).

VPS10 Domain Containing Receptors

The members of the VPS10 domain containing receptor family all contain a domain with homology to the yeast vacuolar sorting protein 10 (VPS10) receptor. This family includes the mosaic receptor SorLA, the neurotensin receptor sortilin, and SorCS, which is expressed during mouse embryonal and early postnatal nervous system development (Hermey, G. et al. (1999) Biochem. Biophys. Res. Commun. 266:347-351; Hermey, G. et al. (2001) Neuroreport 12:29-32). A recently identified member of this family, SorCS2, is highly expressed in the developing and mature mouse central nervous system. Its main site of expression is the floor plate, and high levels are also detected transiently in brain regions including the dopaminergic brain nuclei and the dorsal thalamus (Rezgaoui, M. (2001) Mech. Dev. 100:335-338).

Membrane-Associated Proteins

Tetraspan Family Proteins

The transmembrane 4 superfamily (TM4SF) or tetraspan family is a multigene family encoding type III integral membrane proteins (Wright, M.D. and M.G. Tomlinson (1994) Immunol. Today 15:588-594). The TM4SF is comprised of membrane proteins which traverse the cell membrane four times. Members of the TM4SF include platelet and endothelial cell membrane proteins, melanoma-associated antigens, leukocyte surface glycoproteins, colonal carcinoma antigens, tumor-associated antigens, and surface proteins of the schistosome parasites (Jankowski, S.A. (1994) Oncogene 9:1205-1211). Members of the TM4SF share about 25-30% amino acid sequence identity with one another. A number of TM4SF members have been implicated in signal transduction, control of cell adhesion, regulation of cell growth and proliferation, including development and oncogenesis, and cell motility, including tumor cell metastasis. Expression of TM4SF proteins is associated with a variety of tumors and the level of expression may be altered when cells are growing or activated.

Tumor Antigens

Tumor antigens are surface molecules that are differentially expressed in tumor cells relative to normal cells. Tumor antigens distinguish tumor cells immunologically from normal cells and provide diagnostic and therapeutic targets for human cancers (Takagi, S. et al. (1995) Int. J. Cancer 61:706-715; Liu, E. et al. (1992) Oncogene 7:1027-1032).

Ion Channels

Ion channels are found in the plasma membranes of virtually every cell in the body. For example, chloride channels mediate a variety of cellular functions including regulation of membrane potentials and absorption and secretion of ions across epithelial membranes. When present in intracellular membranes of the Golgi apparatus and endocytic vesicles, chloride channels also regulate organelle pH. (See, e.g., Greger, R. (1988) Annu. Rev. Physiol. 50:111-122.) Electrophysiological and pharmacological properties of chloride channels, including ion conductance, current-voltage relationships, and sensitivity to modulators, suggest that different chloride channels exist in muscles, neurons, fibroblasts, epithelial cells, and lymphocytes. Many channels have sites for phosphorylation by one or more protein kinases including protein kinase A, protein kinase C, tyrosine kinase, and casein kinase II, all of which regulate ion channel activity in cells. Inappropriate phosphorylation of proteins in cells has been linked to changes in cell cycle progression and cell differentiation. Changes in the cell cycle have been linked to induction of apoptosis or cancer. Changes in cell differentiation have been linked to diseases and disorders of the reproductive system, immune system, and skeletal muscle.

Cerebellar granule neurons possess a non-inactivating potassium current which modulates firing frequency upon receptor stimulation by neurotransmitters and controls the resting membrane potential. Potassium channels that exhibit non-inactivating currents include the *ether a go-go* (EAG) channel. A membrane protein designated KCR1 specifically binds to rat EAG by means of its C-terminal region and regulates the cerebellar non-inactivating potassium current. KCR1 is predicted to contain 12 transmembrane domains, with intracellular amino and carboxyl termini. Structural characteristics of these transmembrane regions appear to be similar to those of the transporter superfamily, but no homology between KCR1 and known transporters was found, suggesting that KCR1 belongs to a novel class of transporters. KCR1 appears to be the regulatory component of non-inactivating potassium channels (Hoshi, N. et al. (1998) J. Biol. Chem. 273:23080-23085).

ABC Transporters

ATP-binding cassette (ABC) transporters, also called the "traffic ATPases", are a superfamily of membrane proteins that mediate transport and channel functions in prokaryotes and eukaryotes (Higgins, C.F. (1992) Annu. Rev. Cell Biol. 8:67-113). ABC proteins share a similar overall structure and significant sequence homology. All ABC proteins contain a conserved domain of approximately two hundred amino acid residues which includes one or more nucleotide binding domains. Mutations in ABC transporter genes are associated with various disorders, such as hyperbilirubinemia II/Dubin-Johnson syndrome, recessive Stargardt's disease, X-linked adrenoleukodystrophy, multidrug resistance, celiac disease, and cystic fibrosis.

Semaphorins and Neuropilins

Semaphorins are a large group of axonal guidance molecules consisting of at least 30

different members and are found in vertebrates, invertebrates, and even certain viruses. All semaphorins contain the sema domain which is approximately 500 amino acids in length. Neuropilin, a semaphorin receptor, has been shown to promote neurite outgrowth *in vitro*. The extracellular region of neuropilins consists of three different domains: CUB, discoidin, and MAM domains. The CUB and the MAM motifs of neuropilin have been suggested to have roles in protein-protein interactions and are thought to be involved in the binding of semaphorins through the sema and the C-terminal domains (reviewed in Raper, J.A. (2000) Curr. Opin. Neurobiol. 10:88-94).

Membrane Proteins Associated with Intercellular Communication

Intercellular communication is essential for the development and survival of multicellular organisms. Cells communicate with one another through the secretion and uptake of protein signaling molecules. The uptake of proteins into the cell is achieved by endocytosis, in which the interaction of signaling molecules with the plasma membrane surface, often via binding to specific receptors, results in the formation of plasma membrane-derived vesicles that enclose and transport the molecules into the cytosol. The secretion of proteins from the cell is achieved by exocytosis, in which molecules inside of the cell are packaged into membrane-bound transport vesicles derived from the *trans* Golgi network. These vesicles fuse with the plasma membrane and release their contents into the surrounding extracellular space. Endocytosis and exocytosis result in the removal and addition of plasma membrane components, and the recycling of these components is essential to maintain the integrity, identity, and functionality of both the plasma membrane and internal membrane-bound compartments.

Nogo has been identified as a component of the central nervous system myelin that prevents axonal regeneration in adult vertebrates. Cleavage of the Nogo-66 receptor and other glycophosphatidylinositol-linked proteins from axonal surfaces renders neurons insensitive to Nogo-66, facilitating potential recovery from CNS damage (Fournier, A.E. et al. (2001) Nature 409:341-346).

The slit proteins are extracellular matrix proteins expressed by cells at the ventral midline of the nervous system. Slit proteins are ligands for the repulsive guidance receptor Roundabout (Robo) and thus play a role in repulsive axon guidance (Brose, K. et al. (1999) Cell 96:795-806).

Lysosomes are the site of degradation of intracellular material during autophagy and of extracellular molecules following endocytosis. Lysosomal enzymes are packaged into vesicles which bud from the *trans*-Golgi network. These vesicles fuse with endosomes to form the mature lysosome in which hydrolytic digestion of endocytosed material occurs. Lysosomes can fuse with autophagosomes to form a unique compartment in which the degradation of organelles and other intracellular components occurs.

Protein sorting by transport vesicles, such as the endosome, has important consequences for a

variety of physiological processes including cell surface growth, the biogenesis of distinct intracellular organelles, endocytosis, and the controlled secretion of hormones and neurotransmitters (Rothman, J.E. and F.T. Wieland (1996) Science 272:227-234). In particular, neurodegenerative disorders and other neuronal pathologies are associated with biochemical flaws during endosomal protein sorting or endosomal biogenesis (Mayer, R.J. et al. (1996) Adv. Exp. Med. Biol. 389:261-269).

Peroxisomes are organelles independent from the secretory pathway. They are the site of many peroxide-generating oxidative reactions in the cell. Peroxisomes are unique among eukaryotic organelles in that their size, number, and enzyme content vary depending upon organism, cell type, and metabolic needs (Waterham, H.R. and J.M. Cregg (1997) BioEssays 19:57-66). Genetic defects in peroxisome proteins which result in peroxisomal deficiencies have been linked to a number of human pathologies, including Zellweger syndrome, rhizomelic chonrodysplasia punctata, X-linked adrenoleukodystrophy, acyl-CoA oxidase deficiency, bifunctional enzyme deficiency, classical Refsum's disease, DHAP alkyl transferase deficiency, and acatalasemia (Moser, H.W. and A.B. Moser (1996) Ann. NY Acad. Sci. 804:427-441). In addition, Gartner, J. et al. (1991; Pediatr. Res. 29:141-146) found a 22 kDa integral membrane protein associated with lower density peroxisome-like subcellular fractions in patients with Zellweger syndrome.

Normal embryonic development and control of germ cell maturation is modulated by a number of secretory proteins which interact with their respective membrane-bound receptors. Cell fate during embryonic development is determined by members of the activin/TGF-β superfamily, cadherins, IGF-2, and other morphogens. In addition, proliferation, maturation, and redifferentiation of germ cell and reproductive tissues are regulated, for example, by IGF-2, inhibins, activins, and follistatins (Petraglia, F. (1997) Placenta 18:3-8; Mather, J.P. et al. (1997) Proc. Soc. Exp. Biol. Med. 215:209-222). Transforming growth factor beta (TGFbeta) signal transduction is mediated by two receptor Ser/Thr kinases acting in series, type II TGFbeta receptor and (TbetaR-II) phosphorylating type I TGFbeta receptor (TbetaR-I). TbetaR-I-associated protein-1 (TRECAP-1), which distinguishes between quiescent and activated forms of the type I transforming growth factor beta receptor, has been associated with TGFbeta signaling (Charng, M.J. et al. (1998) J. Biol. Chem. 273:9365-9368).

Retinoic acid receptor alpha (RAR alpha) mediates retinoic-acid induced maturation and has been implicated in myeloid development. Genes induced by retinoic acid during granulocytic differentiation include E3, a hematopoietic-specific gene that is an immediate target for the activated RAR alpha during myelopoiesis (Scott, L.M. et al. (1996) Blood 88:2517-2530).

The μ-opioid receptor (MOR) mediates the actions of analgesic agents including morphine, codeine, methadone, and fentanyl as well as heroin. MOR is functionally coupled to a G-protein-activated potassium channel (Mestek A. et al. (1995) J. Neurosci. 15:2396-2406). A variety of MOR

subtypes exist. Alternative splicing has been observed with MOR-1 as with a number of G protein-coupled receptors including somatostatin 2, dopamine D2, prostaglandin EP3, and serotonin receptor subtypes 5-hydroxytryptamine4 and 5-hydroxytryptamine7 (Pan, Y.X. et al. (1999) Mol. Pharm. 56:396-403).

Peripheral and Anchored Membrane Proteins

Some membrane proteins are not membrane-spanning but are attached to the plasma membrane via membrane anchors or interactions with integral membrane proteins. Membrane anchors are covalently joined to a protein post-translationally and include such moieties as prenyl, myristyl, and glycosylphosphatidyl inositol groups. Membrane localization of peripheral and anchored proteins is important for their function in processes such as receptor-mediated signal transduction. For example, prenylation of Ras is required for its localization to the plasma membrane and for its normal and oncogenic functions in signal transduction.

Expression profiling

Microarrays are analytical tools used in bioanalysis. A microarray has a plurality of molecules spatially distributed over, and stably associated with, the surface of a solid support. Microarrays of polypeptides, polynucleotides, and/or antibodies have been developed and find use in a variety of applications, such as gene sequencing, monitoring gene expression, gene mapping, bacterial identification, drug discovery, and combinatorial chemistry.

One area in particular in which microarrays find use is in gene expression analysis. Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

Adipocytes

Adipocytes are the primary site for whole body energy storage, mainly in the form of triglycerides and fatty acids, and also secrete a variety of signaling molecules. Loss of adipose tissue (lipodystrophies) in animal models and humans leads to metabolic disorders that result in severe insulin resistance and potential diabetes. Adipocyte-derived signals play a central role in the maintenance of energy homeostasis by regulating insulin secretion, insulin action, glucose and lipid metabolism, energy balance, host defense and reproduction. These secretory factors include enzymes such as lipoprotein lipase (LPL) and adipsin, growth factors such as vascular endothelial growth factor (VEGF), cytokines such as tumor necrosis factor-alpha, and interleukin 6, and other hormones

involved in fatty acid and glucose metabolism such as leptin, Acrp30, resistin, acylation stimulation protein, and the GLUT4 glucose transporter (Mora, S., and Pessin, J.E. (2002) Diabetes Metab. Res. Rev. 18:345-356).

Lung Cancer

The potential application of gene expression profiling is particularly relevant to improving diagnosis, prognosis, and treatment of cancer, such as lung cancer. Lung cancer is the leading cause of cancer death in the United States, affecting more than 100,000 men and 50,000 women each year. Nearly 90% of the patients diagnosed with lung cancer are cigarette smokers. Tobacco smoke contains thousands of noxious substances that induce carcinogen metabolizing enzymes and covalent DNA adduct formation in the exposed bronchial epithelium. In nearly 80% of patients diagnosed with lung cancer, metastasis has already occurred. Most commonly lung cancers metastasize to pleura, brain, bone, pericardium, and liver. The decision to treat with surgery, radiation therapy, or chemotherapy is made on the basis of tumor histology, response to growth factors or hormones, and sensitivity to inhibitors or drugs. With current treatments, most patients die within one year of diagnosis. Earlier diagnosis and a systematic approach to identification, staging, and treatment of lung cancer could positively affect patient outcome.

Lung cancers progress through a series of morphologically distinct stages from hyperplasia to invasive carcinoma. Malignant lung cancers are divided into two groups comprising four histopathological classes. The non-small cell lung carcinoma (NSCLC) group includes squamous cell carcinomas, adenocarcinomas, and large cell carcinomas and accounts for about 70% of all lung cancer cases. Adenocarcinomas typically arise in the peripheral airways and often form mucin secreting glands. Squamous cell carcinomas typically arise in proximal airways. The histogenesis of squamous cell carcinomas may be related to chronic inflammation and injury to the bronchial epithelium, leading to squamous metaplasia. The small cell lung carcinoma (SCLC) group accounts for about 20% of lung cancer cases. SCLCs typically arise in proximal airways and exhibit a number of paraneoplastic syndromes including inappropriate production of adrenocorticotropin and anti-diuretic hormone.

Lung cancer cells accumulate numerous genetic lesions, many of which are associated with cytologically visible chromosomal aberrations. The high frequency of chromosomal deletions associated with lung cancer may reflect the role of multiple tumor suppressor loci in the etiology of this disease. Deletion of the short arm of chromosome 3 is found in over 90% of cases and represents one of the earliest genetic lesions leading to lung cancer. Deletions at chromosome arms 9p and 17p are also common. Other frequently observed genetic lesions include overexpression of telomerase, activation of oncogenes such as K-ras and c-myc, and inactivation of tumor suppressor genes such as RB, p53 and CDKN2.

Genes differentially regulated in lung cancer have been identified by a variety of methods. Using mRNA differential display technology, Manda *et al.* (1999; Genomics 51:5-14) identified five genes differentially expressed in lung cancer cell lines compared to normal bronchial epithelial cells. Among the known genes, pulmonary surfactant apoprotein A and alpha 2 macroglobulin were down regulated whereas nm23H1 was upregulated. Petersen *et al.* (2000; Int. J. Cancer, 86:512-517) used suppression subtractive hybridization to identify 552 clones differentially expressed in lung tumor derived cell lines, 205 of which represented known genes. Among the known genes, thrombospondin-1, fibronectin, intercellular adhesion molecule 1, and cytokeratins 6 and 18 were previously observed to be differentially expressed in lung cancers. Wang *et al.* (2000; Oncogene 19:1519-1528) used a combination of microarray analysis and subtractive hybridization to identify 17 genes differentially overexpresssed in squamous cell carcinoma compared with normal lung epithelium. Among the known genes they identified were keratin isoform 6, KOC, SPRC, IGFb2, connexin 26, plakofillin 1 and cytokeratin 13.

Prostate cancer

Prostate cancer is a common malignancy in men over the age of 50, and the incidence increases with age. In the US, there are approximately 132,000 newly diagnosed cases of prostate cancer and more than 33,000 deaths from the disorder each year.

Once cancer cells arise in the prostate, they are stimulated by testosterone to a more rapid growth. Thus, removal of the testes can indirectly reduce both rapid growth and metastasis of the cancer. Over 95 percent of prostatic cancers are adenocarcinomas which originate in the prostatic acini. The remaining 5 percent are divided between squamous cell and transitional cell carcinomas, both of which arise in the prostatic ducts or other parts of the prostate gland.

As with most tumors, prostate cancer develops through a multistage progression ultimately resulting in an aggressive tumor phenotype. The initial step in tumor progression involves the hyperproliferation of normal luminal and/or basal epithelial cells. Androgen responsive cells become hyperplastic and evolve into early-stage tumors. Although early-stage tumors are often androgen sensitive and respond to androgen ablation, a population of androgen independent cells evolve from the hyperplastic population. These cells represent a more advanced form of prostate tumor that may become invasive and potentially become metastatic to the bone, brain, or lung. A variety of genes may be differentially expressed during tumor progression. For example, loss of heterozygosity (LOH) is frequently observed on chromosome 8p in prostate cancer. Fluorescence *in situ* hybridization (FISH) revealed a deletion for at least 1 locus on 8p in 29 (69%) tumors, with a significantly higher frequency of the deletion on 8p21.2-p21.1 in advanced prostate cancer than in localized prostate cancer, implying that deletions on 8p22-p21.3 play an important role in tumor differentiation, while 8p21.2-p21.1 deletion plays a role in progression of prostate cancer (Oba, K. et al. (2001) Cancer

Genet. Cytogenet. 124: 20-26).

A primary diagnostic marker for prostate cancer is prostate specific antigen (PSA). PSA is a tissue-specific serine protease almost exclusively produced by prostatic epithelial cells. The quantity of PSA correlates with the number and volume of the prostatic epithelial cells, and consequently, the levels of PSA are an excellent indicator of abnormal prostate growth. Men with prostate cancer exhibit an early linear increase in PSA levels followed by an exponential increase prior to diagnosis. However, since PSA levels are also influenced by factors such as inflammation, androgen and other growth factors, some scientists maintain that changes in PSA levels are not useful in detecting individual cases of prostate cancer.

Current areas of cancer research provide additional prospects for markers as well as potential therapeutic targets for prostate cancer. Several growth factors have been shown to play a critical role in tumor development, growth, and progression. The growth factors Epidermal Growth Factor (EGF), Fibroblast Growth Factor (FGF), and Tumor Growth Factor alpha (TGFα) are important in the growth of normal as well as hyperproliferative prostate epithelial cells, particularly at early stages of tumor development and progression, and affect signaling pathways in these cells in various ways (Lin, J. et al. (1999) Cancer Res. 59:2891-2897; Putz, T. et al. (1999) Cancer Res. 59:227-233). The TGF-β family of growth factors are generally expressed at increased levels in human cancers and the high expression levels in many cases correlates with advanced stages of malignancy and poor survival (Gold, L.I. (1999) Crit. Rev. Oncog. 10:303-360). Finally, there are human cell lines representing both the androgen-dependent stage of prostate cancer (LNCap) as well as the androgen-independent, hormone refractory stage of the disease (PC3 and DU-145) that have proved useful in studying gene expression patterns associated with the progression of prostate cancer, and the effects of cell treatments on these expressed genes (Chung, T.D. (1999) Prostate 15:199-207).

Ovarian cancer

Ovarian cancer is the leading cause of death from a gynecologic cancer. The majority of ovarian cancers are derived from epithelial cells, and 70% of patients with epithelial ovarian cancers present with late-stage disease. As a result, the long-term survival rate for this disease is very low. Identification of early-stage markers for ovarian cancer would significantly increase the survival rate. Genetic variations involved in ovarian cancer development include mutation of p53 and microsatellite instability. Gene expression patterns likely vary when normal ovary is compared to ovarian tumors.

Breast cancer

More than 180,000 new cases of breast cancer are diagnosed each year, and the mortality rate for breast cancer approaches 10% of all deaths in females between the ages of 45-54 (Gish, K. (1999) AWIS Magazine 28:7-10). However the survival rate based on early diagnosis of localized breast cancer is extremely high (97%), compared with the advanced stage of the disease in which the tumor

has spread beyond the breast (22%). Current procedures for clinical breast examination are lacking in sensitivity and specificity, and efforts are underway to develop comprehensive gene expression profiles for breast cancer that may be used in conjunction with conventional screening methods to improve diagnosis and prognosis of this disease (Perou, C.M. et al. (2000) Nature 406:747-752).

Mutations in two genes, BRCA1 and BRCA2, are known to greatly predispose a woman to breast cancer and may be passed on from parents to children (Gish, *supra*). However, this type of hereditary breast cancer accounts for only about 5% to 9% of breast cancers, while the vast majority of breast cancer is due to non-inherited mutations that occur in breast epithelial cells.

The relationship between expression of epidermal growth factor (EGF) and its receptor, EGFR, to human mammary carcinoma has been particularly well studied (Khazaie, K. *et al.* (1993) Cancer and Metastasis Rev. 12:255-274, and references cited therein). Overexpression of EGFR, particularly coupled with down-regulation of the estrogen receptor, is a marker of poor prognosis in breast cancer patients. In addition, EGFR expression in breast tumor metastases is frequently elevated relative to the primary tumor, suggesting that EGFR is involved in tumor progression and metastasis. This is supported by accumulating evidence that EGF has effects on cell functions related to metastatic potential, such as cell motility, chemotaxis, secretion and differentiation. Changes in expression of other members of the erbB receptor family, of which EGFR is one, have also been implicated in breast cancer. The abundance of erbB receptors, such as HER-2/neu, HER-3, and HER-4, and their ligands in breast cancer points to their functional importance in the pathogenesis of the disease, and may therefore provide targets for therapy of the disease (Bacus, S.S. *et al.* (1994) Am. J. Clin. Pathol. 102:S13-S24). Other known markers of breast cancer include a human secreted frizzled protein mRNA that is downregulated in breast tumors; the matrix G1a protein which is overexpressed in human breast carcinoma cells; Drg1 or RTP, a gene whose expression is diminished in colon, breast, and prostate tumors; maspin, a tumor suppressor gene downregulated in invasive breast carcinomas; and CaN19, a member of the S100 protein family, all of which are down-regulated in mammary carcinoma cells relative to normal mammary epithelial cells (Zhou, Z. *et al.* (1998) Int. J. Cancer 78:95-99; Chen, L. *et al.* (1990) Oncogene 5:1391-1395; Ulrix, W. *et al.* (1999) FEBS Lett. 455:23-26; Sager, R. *et al.* (1996) Curr. Top. Microbiol. Immunol. 213:51-64; and Lee, S.W. *et al.* (1992) Proc. Natl. Acad. Sci. USA 89:2504-2508).

Cell lines derived from human mammary epithelial cells at various stages of breast cancer provide a useful model to study the process of malignant transformation and tumor progression as it has been shown that these cell lines retain many of the properties of their parental tumors for lengthy culture periods (Wistuba, I.I. *et al.* (1998) Clin. Cancer Res. 4:2931-2938). Such a model is particularly useful for comparing phenotypic and molecular characteristics of human mammary epithelial cells at various stages of malignant transformation.

Colon cancer

While soft tissue sarcomas are relatively rare, more than 50% of new patients diagnosed with the disease will die from it. The molecular pathways leading to the development of sarcomas are relatively unknown, due to the rarity of the disease and variation in pathology. Colon cancer evolves through a multi-step process whereby pre-malignant colonocytes undergo a relatively defined sequence of events leading to tumor formation. Several factors participate in the process of tumor progression and malignant transformation including genetic factors, mutations, and selection.

To understand the nature of gene alterations in colorectal cancer, a number of studies have focused on the inherited syndromes. Familial adenomatous polyposis (FAP), is caused by mutations in the adenomatous polyposis coli gene (APC), resulting in truncated or inactive forms of the protein. This tumor suppressor gene has been mapped to chromosome 5q. Hereditary nonpolyposis colorectal cancer (HNPCC) is caused by mutations in mis-match repair genes. Although hereditary colon cancer syndromes occur in a small percentage of the population and most colorectal cancers are considered sporadic, knowledge from studies of the hereditary syndromes can be generally applied. For instance, somatic mutations in APC occur in at least 80% of sporadic colon tumors. APC mutations are thought to be the initiating event in the disease. Other mutations occur subsequently. Approximately 50% of colorectal cancers contain activating mutations in ras, while 85% contain inactivating mutations in p53. Changes in all of these genes lead to gene expression changes in colon cancer.

Neurological disorders

Characterization of region-specific gene expression in the human brain provides a context and background for molecular neurobiology on a variety of neurological disorders. For example, Alzheimer's disease (AD) is a progressive, neurodestructive process of the human neocortex, characterized by the deterioration of memory and higher cognitive function. A progressive and irreversible brain disorder, AD is characterized by three major pathogenic episodes involving (a) an aberrant processing and deposition of beta-amyloid precursor protein (betaAPP) to form neurotoxic beta-amyloid (betaA) peptides and an aggregated insoluble polymer of betaA that forms the senile plaque, (b) the establishment of intraneuronal neuritic tau pathology yielding widespread deposits of agyrophilic neurofibrillary tangles (NFT) and (c) the initiation and proliferation of a brain-specific inflammatory response. These three seemingly disperse attributes of AD etiopathogenesis are linked by the fact that proinflammatory microglia, reactive astrocytes and their associated cytokines and chemokines are associated with the biology of the microtubule associated protein tau, betaA speciation and aggregation. Missense mutations in the presenilin genes PS1 and PS2, implicated in early onset familial AD, cause abnormal betaAPP processing with resultant overproduction of betaA42 and related neurotoxic peptides. Specific betaA fragments such as betaA42 can further potentiate proinflammatory mechanisms. Expression of the inducible oxidoreductase

cyclooxygenase-2 and cytosolic phospholipase A2 (cPLA2) is strongly activated during cerebral ischemia and trauma, epilepsy and AD, indicating the induction of proinflammatory gene pathways as a response to brain injury. Neurotoxic metals such as aluminum and zinc, both implicated in AD etiopathogenesis, and arachidonic acid, a major metabolite of brain cPLA2 activity, each polymerize hyperphosphorylated tau to form NFT-like bundles. Studies have identified a reduced risk for AD in patients aged over 70 years who were previously treated with non-steroidal anti-inflammatory drugs for non-CNS afflictions that include arthritis. (For a review of the interrelationships between the mechanisms of PS1, PS2 and betaAPP gene expression, tau and betaA deposition and the induction, regulation and proliferation in AD of the neuroinflammatory response, see Lukiw, W.J and Bazan, N.G. (2000) Neurochem. Res. 2000 25:1173-1184).

Osteosarcoma

Osteosarcoma is the most common malignant bone tumor in children. Approximately 80% of patients present with non-metastatic disease. After the diagnosis is made by an initial biopsy, treatment involves the use of 3-4 courses of neoadjuvant chemotherapy before definitive surgery, followed by post-operative chemotherapy. With currently available treatment regimens, approximately 30-40% of patients with non-metastatic disease relapse after therapy. Currently, there is no prognostic factor that can be used at the time of initial diagnosis to predict which patients will have a high risk of relapse. The only significant prognostic factor predicting the outcome in a patient with non-metastatic osteosarcoma is the histopathologic response of the primary tumor resected at the time of definitive surgery. The degree of necrosis in the primary tumor is a reflection of the tumor response to neoadjuvant chemotherapy. A higher degree of necrosis (good or favorable response) is associated with a lower risk of relapse and a better outcome. Patients with a lower degree of necrosis (poor or unfavorable response) have a much higher risk of relapse and poor outcome even after complete resection of the primary tumor. Unfortunately, poor outcome cannot be altered despite modification of post-operative chemotherapy to account for the resistance of the primary tumor to neoadjuvant chemotherapy. Thus, there is an urgent need to identify prognostic factors that can be used at the time of diagnosis to recognize the subtypes of osteosarcomas that have various risks of relapse, so that more appropriate chemotherapy can be used at the outset to improve the outcome.

Inflammation and Immune Response

Human peripheral blood mononuclear cells (PBMCs) represent the major cellular components of the immune system. PBMCs contain about 12% B lymphocytes, 25% CD4+ and 15% CD8+ lymphocytes, 20% NK cells, 25% monocytes, and 3% various cells that include dendritic cells and progenitor cells. The proportions, as well as the biology of these cellular components tend to vary slightly between healthy individuals, depending on factors such as age, gender, past medical history, and genetic background.

T cells can be subdivided into two classes according to their main function and the surface antigens they express. First, CD4 positive (+) T cells, also known as T Helper cells, primarily regulate the immune response by producing soluble factors that, in turn, regulate the activity of effector cells such as B lymphocytes, NK cells, and macrophages. Second, CD8 positive (+) T cells, also known as cytotoxic T cells, primarily kill "abnormal" cells such as tumor cells or cells infected by viruses. In the blood of a healthy adult, CD4+ T cells and CD8+ T cells represent 25% and 15% of the mononuclear cells, respectively. These two T cell populations can be readily expanded out of blood by incubating bulk peripheral blood mononuclear cells (PBMCs) in the presence of phytohemagglutinin (PHA) and interleukin 2 (IL-2). After 8 to 10 days of treatment, both CD4+ and CD8+ T cells expand roughly 5- to 10-fold, yielding a cell population composed of >90% T cells, also known as PHA blasts. T cell expansion occurs during the first 5 days of PHA stimulation; after 8 to 10 days in culture, most PHA blasts have returned to a resting state.

T cells require two distinct signals to achieve optimal activation. First, the "antigenic" signal delivered through the binding of the TCR-CD3 complex. Second, the costimulatory signal delivered through the binding of the CD28 molecules. Upon binding of the TCR-CD3 complex alone, T cells only achieve a partial state of activation. However, it is important to note that the signaling requirements of T cells depend greatly on the cycling state of those cells.

Staphylococcal exotoxins such as staphlococcal exotoxin B (SEB) specifically activate human T cells, expressing an appropriate TCR-Vβ chain. Although polyclonal in nature, T cells activated by Staphylococcal exotoxins require antigen presenting cells (APCs) to present the exotoxin molecules to the T cells and deliver the costimulatory signals required for optimum T cell activation. Although Staphylococcal exotoxins must be presented to T cells by APCs, these molecules are not required to be processed by APC. Indeed, Staphylococcal exotoxins directly bind to a non-polymorphic portion of the human MHC class II molecules, bypassing the need for capture, cleavage, and binding of the peptides to the polymorphic antigenic groove of the MHC class II molecules.

PBMCs are used as a model system to study the changes in gene expression resulting from activation of cells of the immune system. PMA is a broad activator of the protein kinase C-dependent pathways. Ionomycin is a calcium ionophore that permits the entry of calcium in the cell, hence increasing the cytosolic calcium concentration. The combination of PMA and ionomycin activates two of the major signaling pathways used by mammalian cells to interact with their environment. In PBMCs, the combination of PMA and ionomycin mimics the secondary signaling events elicited during activation of lymphocytes, NK cells, and monocytes.

Beclomethasone is a synthetic glucocorticoid that is used for treating steroid-dependent asthma, relieving symptoms associated with allergic or nonallergic (vasomotor) rhinitis, or preventing recurrent nasal polyps following surgical removal. The anti-inflammatory and vasoconstrictive

effects of intranasal beclomethasone are 5000 times greater than those produced by hydrocortisone. Glucocorticoids are naturally occurring hormones that prevent or suppress inflammation and immune responses when administered at pharmacological doses. At the molecular level, unbound glucocorticoids readily cross cell membranes and bind with high affinity to specific cytoplasmic receptors. Subsequent to binding, transcription and, ultimately, protein synthesis are affected. The result can include inhibition of leukocyte infiltration at the site of inflammation, interference in the function of mediators of inflammatory response, and suppression of humoral immune responses. The antiinflammatory actions of corticosteroids are thought to involve phospholipase A2 inhibitory proteins, collectively called lipocortins. Lipocortins, in turn, control the biosynthesis of potent mediators of inflammation such as prostaglandins and leukotrienes by inhibiting the release of the precursor molecule arachidonic acid.

Cytokines

Interleukin 1 beta (IL-1b) is a cytokine associated with acute inflammatory responses and is generally considered the prototypical pro-inflammatory cytokine. However, IL-1β functions are not limited to the inflammatory response since this molecule is involved in processes such as fever induction, metabolic regulation, and bone remodeling. Both cells of the immune system (monocytes, dendritic cells, NK cells, platelets, and neutrophils) and somatic cells (osteoblasts, neurons, Schwann's cells, oligodendrocytes, and adrenal cortical cells) can produce IL-1β. IL-1β has been shown to induce its own production in monocytes; induce the production of adhesion molecules and chemokines in endothelial cells; and in conjunction with IL-12, induce interferon-γ production by NK Cells. IL-1β is produced as a single chain pro-molecule that needs to be cleaved by a specialized protease, IL-1β Converting Enzyme (ICE), to acquire its function.

Interleukin 5 (IL-5) is a T cell-derived factor that promotes the proliferation, differentiation, and activation of eosinophils. IL-5 has also been known as T cell replacing factor (TRF), B cell growth factor II (BCGFII), B cell differentiation factor m (BCDF m), eosinophil differentiation factor (EDF), and eosinophil colony-stimulating factor (Eo-CSF). IL-5 exerts its activity on target cells by binding to specific cell surface receptors. The functional high-affinity receptor for human IL-5 is composed of a low-affinity IL-5 binding α-subunit and a non-binding common β-subunit that is shared with the high-affinity receptors for GM-CSF and IL-3.

Interleukin 6 (IL-6) is a multifunctional protein that plays important roles in host defense, acute phase reactions, immune responses, and hematopoiesis. According to the type of biological responses being studied, IL-6 was previously named interferon-b2, 26-kDa protein, B cell stimulatory factor-2 (BSF-2), hybridoma/plasmacytoma growth factor, hepatocyte stimulating factor, cytotoxic T cell differentiation factor, and macrophage-granulocyte inducing factor 2A (MGI-2A). The IL-6 designation was adopted after these variously named proteins were found to be identical on the basis

of their amino acid and/or nucleotide sequences. IL-6 is expressed by a variety of normal and transformed cells including T cells, B cells, monocytes/macrophages, fibroblasts, hepatocytes, keratinocytes, astrocytes, vascular endothelial cells, and various tumor cells. The production of IL-6 is upregulated by numerous signals including mitogenic or antigenic stimulation, LPS, calcium ionophore, IL-1, IL-2, IFN, TNF, PDGF, and viruses. IL-4 and IL-13 inhibit IL-6 expression in monocytes.

There is a need in the art for new compositions, including nucleic acids and proteins, for the diagnosis, prevention, and treatment of cell proliferative, autoimmune/inflammatory, neurological, metabolic, developmental, and endocrine disorders.

SUMMARY OF THE INVENTION

Various embodiments of the invention provide purified polypeptides, receptors and membrane-associated proteins, referred to collectively as 'REMAP' and individually as 'REMAP-1,' 'REMAP-2,' 'REMAP-3,' 'REMAP-4,' 'REMAP-5,' 'REMAP-6,' 'REMAP-7,' 'REMAP-8,' 'REMAP-9,' 'REMAP-10,' 'REMAP-11,' 'REMAP-12,' 'REMAP-13,' 'REMAP-14,' 'REMAP-15,' 'REMAP-16,' 'REMAP-17,' 'REMAP-18,' 'REMAP-19,' 'REMAP-20,' 'REMAP-21,' 'REMAP-22,' 'REMAP-23,' 'REMAP-24,' 'REMAP-25,' 'REMAP-26,' 'REMAP-27,' 'REMAP-28,' 'REMAP-29,' 'REMAP-30,' 'REMAP-31,' 'REMAP-32,' 'REMAP-33,' 'REMAP-34,' 'REMAP-35,' 'REMAP-36,' 'REMAP-37,' 'REMAP-38,' 'REMAP-39,' 'REMAP-40,' 'REMAP-41,' 'REMAP-42,' 'REMAP-43,' 'REMAP-44,' 'REMAP-45,' 'REMAP-46,' 'REMAP-47,' 'REMAP-48,' and 'REMAP-49,' and methods for using these proteins and their encoding polynucleotides for the detection, diagnosis, and treatment of diseases and medical conditions. Embodiments also provide methods for utilizing the purified receptors and membrane-associated proteins and/or their encoding polynucleotides for facilitating the drug discovery process, including determination of efficacy, dosage, toxicity, and pharmacology. Related embodiments provide methods for utilizing the purified receptors and membrane-associated proteins and/or their encoding polynucleotides for investigating the pathogenesis of diseases and medical conditions.

An embodiment provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-49. Another embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID

NO:1-49.

Still another embodiment provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-49. In another embodiment, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-49. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:50-98.

Still another embodiment provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-49. Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another embodiment provides a transgenic organism comprising the recombinant polynucleotide.

Another embodiment provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-49. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Yet another embodiment provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, b) a polypeptide comprising a

naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-49.

Still yet another embodiment provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:50-98, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:50-98, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In other embodiments, the polynucleotide can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:50-98, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:50-98, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex. In a related embodiment, the method can include detecting the amount of the hybridization complex. In still other embodiments, the probe can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Still yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:50-98, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:50-98, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain

reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof. In a related embodiment, the method can include detecting the amount of the amplified target polynucleotide or fragment thereof.

Another embodiment provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, and a pharmaceutically acceptable excipient. In one embodiment, the composition can comprise an amino acid sequence selected from the group consisting of SEQ ID NO:1-49. Other embodiments provide a method of treating a disease or condition associated with decreased or abnormal expression of functional REMAP, comprising administering to a patient in need of such treatment the composition.

Yet another embodiment provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-49. The method comprises a) contacting a sample comprising the polypeptide with a compound, and b) detecting agonist activity in the sample. Another embodiment provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with decreased expression of functional REMAP, comprising administering to a patient in need of such treatment the composition.

Still yet another embodiment provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, and d) an immunogenic fragment of a polypeptide having an amino

acid sequence selected from the group consisting of SEQ ID NO:1-49. The method comprises a) contacting a sample comprising the polypeptide with a compound, and b) detecting antagonist activity in the sample. Another embodiment provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with overexpression of functional REMAP, comprising administering to a patient in need of such treatment the composition.

Another embodiment provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-49. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

Yet another embodiment provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-49, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-49. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

Still yet another embodiment provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:50-98, the method comprising a) contacting a sample comprising the target polynucleotide with a compound, b)

detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

Another embodiment provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:50-98, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:50-98, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:50-98, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:50-98, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for full length polynucleotide and polypeptide embodiments of the invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptide embodiments of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide embodiments, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide embodiments, along with selected fragments of the polynucleotides.

Table 5 shows representative cDNA libraries for polynucleotide embodiments.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze polynucleotides and polypeptides, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide sequences of the invention, along with allele frequencies in different human populations.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleic acids, and methods are described, it is understood that embodiments of the invention are not limited to the particular machines, instruments, materials, and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the invention.

As used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with various embodiments of the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"REMAP" refers to the amino acid sequences of substantially purified REMAP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of REMAP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other

compound or composition which modulates the activity of REMAP either by directly interacting with REMAP or by acting on components of the biological pathway in which REMAP participates.

An "allelic variant" is an alternative form of the gene encoding REMAP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding REMAP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as REMAP or a polypeptide with at least one functional characteristic of REMAP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding REMAP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide encoding REMAP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent REMAP. Deliberate amino acid substitutions may be made on the basis of one or more similarities in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of REMAP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" can refer to an oligopeptide, a peptide, a polypeptide, or a protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid. Amplification may be carried out using polymerase chain reaction (PCR) technologies or other nucleic acid amplification technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity

of REMAP. Antagonists may include proteins such as antibodies, anticalins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of REMAP either by directly interacting with REMAP or by acting on components of the biological pathway in which REMAP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind REMAP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker (Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13).

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-

handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a polynucleotide having a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic REMAP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide" and a "composition comprising a given polypeptide" can refer to any composition containing the given polynucleotide or polypeptide. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotides encoding REMAP or fragments of REMAP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (Accelrys,

Burlington MA) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of REMAP or a polynucleotide encoding REMAP which can be identical in sequence to, but shorter in length than, the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from about 5 to about 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:50-98 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:50-98, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:50-98 can be employed in one or more embodiments of methods of the invention, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:50-98 from related polynucleotides. The precise length of a fragment of SEQ ID NO:50-98 and the region of SEQ ID NO:50-98 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-49 is encoded by a fragment of SEQ ID NO:50-98. A fragment of SEQ ID NO:1-49 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-49. For example, a fragment of SEQ ID NO:1-49 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-49. The precise length of a fragment of SEQ ID NO:1-49 and the region of SEQ ID NO:1-49 to which the fragment corresponds can be determined based on the intended purpose for the fragment using one or more analytical methods described herein or otherwise known in the art.

A "full length" polynucleotide is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, alternatively, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of identical nucleotide matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using one or more computer algorithms or programs known in the art or described herein. For example, percent identity can be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989; CABIOS 5:151-153) and in Higgins, D.G. et al. (1992; CABIOS 8:189-191). For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms which can be used is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of identical residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide. The phrases "percent similarity" and "% similarity," as applied to polypeptide sequences, refer to the percentage of residue matches, including identical residue matches and conservative substitutions, between at least two polypeptide sequences aligned using a standardized algorithm. In contrast, conservative substitutions are not included in the calculation of percent identity between polypeptide sequences.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version

2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature

under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY, ch. 9).

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acids by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid present in solution and another nucleic acid immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or polynucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of REMAP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of REMAP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, antibodies, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, antibody, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of REMAP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of REMAP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an REMAP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of REMAP.

"Probe" refers to nucleic acids encoding REMAP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acids. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also

be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in, for example, Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY), Ausubel, F.M. et al. (1999; Short Protocols in Molecular Biology, 4th ed., John Wiley & Sons, New York NY), and Innis, M. et al. (1990; PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a nucleic acid that is not naturally occurring or has a

sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook and Russell (*supra*). The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA molecule, is composed of the same linear sequence of nucleotides as the reference DNA molecule with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing REMAP, nucleic acids encoding REMAP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are

removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably at least about 75% free, and most preferably at least about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In another embodiment, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook and Russell (*supra*).

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having

at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotides that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity or sequence similarity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity or sequence similarity over a certain defined length of one of the polypeptides.

THE INVENTION

Various embodiments of the invention include new human receptors and membrane-associated proteins (REMAP), the polynucleotides encoding REMAP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, autoimmune/inflammatory, neurological, metabolic, developmental, and endocrine disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an

Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown. Column 6 shows the Incyte ID numbers of physical, full length clones corresponding to the polypeptide and polynucleotide sequences of the invention. The full length clones encode polypeptides which have at least 95% sequence identity to the polypeptide sequences shown in column 3.

Table 2 shows sequences with homology to polypeptide embodiments of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows amino acid residues comprising signature sequences, domains, motifs, potential phosphorylation sites, and potential glycosylation sites. Column 5 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are receptors and membrane-associated proteins. For example, SEQ ID NO:3 is 92% identical, from residue S91 to residue K379, to human CD33L1 antigen (GenBank ID g2913995) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 5.8E-155, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:3 also has homology to sialic acid binding Ig-like lectin 6 (CD33 antigen-like), as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:3 also contains an immunoglobulin domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM and SMART databases of conserved protein families/domains, an immunoglobulin C-2 type domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-

based SMART database of conserved protein families/domains, and an IG superfamily domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based INCY database of conserved protein families/domains. Data from BLIMPS and MOTIFS analyses, and BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:3 is an immunoglobulin-like lectin. In another example, SEQ ID NO:15 is a splice variant of human endoglin (GenBank ID g402207) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.5e-288, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:15 also has homology to proteins that are localized to the plasma membrane, bind transforming growth factor beta (TGF-β), and play a critical role in angiogenesis and heart development, as determined by BLAST analysis using the PROTEOME database. BLAST analyses against the PRODOM and DOMO databases provide further corroborative evidence that SEQ ID NO:15 is a membrane-associated receptor for TGF-β. In another example, SEQ ID NO:22 is 99% identical, from residue M1 to residue T236, to human nuclear orphan receptor LXR-alpha (GenBank ID g726513) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.1E-126, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:22 also has homology to nuclear receptor subfamily 1 group H member 3, a ligand-dependent nuclear receptor transcription factor, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:22 also contains a C4 zinc finger domain and a C4 zinc finger in nuclear hormone receptor domain, as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM and SMART databases of conserved protein families/domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses, and BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:22 is a nuclear receptor LXR-alpha. In another example, SEQ ID NO:36 is a splice variant of human thyroid hormone receptor (GenBank ID g339665) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 9.9e-244, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:36 also has homology to proteins that are localized to the nucleus, bind DNA, and act as ligand-activated transcription factors, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:36 also contains a ligand-binding domain of nuclear hormone receptors, and a C4-type zinc finger found in nuclear hormone receptors, as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM and SMART databases of conserved protein families/domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses, and BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID

NO:36 is a nuclear hormone receptor. In another example, SEQ ID NO:40 (525 amino acids in length) is 100% identical, from residue M1 to residue E471, to human Eph-family protein (GenBank ID g2281008) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.3e-266, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:40 also has homology to proteins that are localized to the plasma membrane, have kinase and signaling activity, that bind ephrin B2, that play a role in T cell development and apoptosis, and that are associated with decreased tumorigenicity when upregulated, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:40 also contains a fibronectin type 3 domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM and SMART databases of conserved protein families/domains, as well as an Ephrin receptor ligand binding domain, as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein families/domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses, and BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:40 is a member of the Eph receptor family of proteins. In another example, SEQ ID NO:46 is 100% identical, from residue M1 to residue N773, to human CSF-1 receptor (GenBank ID g1915976) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:46 also has homology to proteins that are localized to the plasma membrane, function as protein kinases, and are colony stimulating factor 1 receptors, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:46 also contains immunoglobulin domains, immunoglobulin C-2 type domains, and a tyrosine kinase, catalytic domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based SMART database of conserved protein families/domains; immunoglobulin domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein families/domains; and Ig superfamily from SCOP domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based INCY database of conserved protein families/domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses, and BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:46 is a receptor tyrosine kinase. SEQ ID NO:1-2, SEQ ID NO:4-14, SEQ ID NO:16-21, SEQ ID NO:23-35, SEQ ID NO:37-39, SEQ ID NO:41-45, and SEQ ID NO:47-49 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-49 are described in Table 7.

As shown in Table 4, the full length polynucleotide embodiments were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of

these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide embodiments, and of fragments of the polynucleotides which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:50-98 or that distinguish between SEQ ID NO:50-98 and related polynucleotides.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotides. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The

following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotides which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotides. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows single nucleotide polymorphisms (SNPs) found in polynucleotide sequences of the invention, along with allele frequencies in different human populations. Columns 1 and 2 show the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence. Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected) indicates that the frequency of allele 1 in the population was too low to be detected, while n/a (not available) indicates that the allele frequency was not determined for the population.

The invention also encompasses REMAP variants. Various embodiments of REMAP variants can have at least about 80%, at least about 90%, or at least about 95% amino acid sequence

identity to the REMAP amino acid sequence, and can contain at least one functional or structural characteristic of REMAP.

Various embodiments also encompass polynucleotides which encode REMAP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:50-98, which encodes REMAP. The polynucleotide sequences of SEQ ID NO:50-98, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses variants of a polynucleotide encoding REMAP. In particular, such a variant polynucleotide will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a polynucleotide encoding REMAP. A particular aspect of the invention encompasses a variant of a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NO:50-98 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:50-98. Any one of the polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of REMAP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide encoding REMAP. A splice variant may have portions which have significant sequence identity to a polynucleotide encoding REMAP, but will generally have a greater or lesser number of nucleotides due to additions or deletions of blocks of sequence arising from alternate splicing during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to a polynucleotide encoding REMAP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide encoding REMAP. For example, a polynucleotide comprising a sequence of SEQ ID NO:51, a polynucleotide comprising a sequence of SEQ ID NO:60, a polynucleotide comprising a sequence of SEQ ID NO:62, and a polynucleotide comprising a sequence of SEQ ID NO:63 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:54, a polynucleotide comprising a sequence of SEQ ID NO:57, a polynucleotide comprising a sequence of SEQ ID NO:58, and a polynucleotide comprising a sequence of SEQ ID NO:59 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:55 and a polynucleotide comprising a sequence of SEQ ID NO:56 are splice variants of each other; and a polynucleotide comprising a sequence of SEQ ID NO:67 and a polynucleotide comprising a sequence of SEQ ID NO:89 are splice variants of each

other. Any one of the splice variants described above can encode a polypeptide which contains at least one functional or structural characteristic of REMAP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding REMAP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring REMAP, and all such variations are to be considered as being specifically disclosed.

Although polynucleotides which encode REMAP and its variants are generally capable of hybridizing to polynucleotides encoding naturally occurring REMAP under appropriately selected conditions of stringency, it may be advantageous to produce polynucleotides encoding REMAP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding REMAP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of polynucleotides which encode REMAP and REMAP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic polynucleotide may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a polynucleotide encoding REMAP or any fragment thereof.

Embodiments of the invention can also include polynucleotides that are capable of hybridizing to the claimed polynucleotides, and, in particular, to those having the sequences shown in SEQ ID NO:50-98 and fragments thereof, under various conditions of stringency (Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations

of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (Ausubel et al., *supra*, ch. 7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

The nucleic acids encoding REMAP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (BD Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze

the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotides or fragments thereof which encode REMAP may be cloned in recombinant DNA molecules that direct expression of REMAP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other polynucleotides which encode substantially the same or a functionally equivalent polypeptides may be produced and used to express REMAP.

The polynucleotides of the invention can be engineered using methods generally known in the art in order to alter REMAP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of REMAP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable

manner.

In another embodiment, polynucleotides encoding REMAP may be synthesized, in whole or in part, using one or more chemical methods well known in the art (Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232). Alternatively, REMAP itself or a fragment thereof may be synthesized using chemical methods known in the art. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of REMAP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography (Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421). The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing (Creighton, *supra*, pp. 28-53).

In order to express a biologically active REMAP, the polynucleotides encoding REMAP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotides encoding REMAP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of polynucleotides encoding REMAP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a polynucleotide sequence encoding REMAP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

Methods which are well known to those skilled in the art may be used to construct expression vectors containing polynucleotides encoding REMAP and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques,

synthetic techniques, and *in vivo* genetic recombination (Sambrook and Russell, *supra*, ch. 1-4, and 8; Ausubel et al., *supra*, ch. 1, 3, and 15).

A variety of expression vector/host systems may be utilized to contain and express polynucleotides encoding REMAP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook and Russell, *supra*; Ausubel et al., *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355). Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of polynucleotides to the targeted organ, tissue, or cell population (Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5:350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:6340-6344; Buller, R.M. et al. (1985) Nature 317:813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31:219-226; Verma, I.M. and N. Somia (1997) Nature 389:239-242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotides encoding REMAP. For example, routine cloning, subcloning, and propagation of polynucleotides encoding REMAP can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen). Ligation of polynucleotides encoding REMAP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509). When large quantities of REMAP are needed, e.g. for the production of antibodies, vectors which direct high level expression of REMAP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of REMAP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such

vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign polynucleotide sequences into the host genome for stable propagation (Ausubel et al., *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184).

Plant systems may also be used for expression of REMAP. Transcription of polynucleotides encoding REMAP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection (The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196).

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, polynucleotides encoding REMAP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses REMAP in host cells (Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355).

For long term production of recombinant proteins in mammalian systems, stable expression of REMAP in cell lines is preferred. For example, polynucleotides encoding REMAP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These

include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; BD Clontech), β-glucuronidase and its substrate β-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding REMAP is inserted within a marker gene sequence, transformed cells containing polynucleotides encoding REMAP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding REMAP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the polynucleotide encoding REMAP and that express REMAP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of REMAP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on REMAP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa

NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding REMAP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, polynucleotides encoding REMAP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with polynucleotides encoding REMAP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode REMAP may be designed to contain signal sequences which direct secretion of REMAP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted polynucleotides or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding REMAP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric REMAP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of REMAP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST),

maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the REMAP encoding sequence and the heterologous protein sequence, so that REMAP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In another embodiment, synthesis of radiolabeled REMAP may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

REMAP, fragments of REMAP, or variants of REMAP may be used to screen for compounds that specifically bind to REMAP. One or more test compounds may be screened for specific binding to REMAP. In various embodiments, 1, 2, 3, 4, 5, 10, 20, 50, 100, or 200 test compounds can be screened for specific binding to REMAP. Examples of test compounds can include antibodies, anticalins, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules.

In related embodiments, variants of REMAP can be used to screen for binding of test compounds, such as antibodies, to REMAP, a variant of REMAP, or a combination of REMAP and/or one or more variants REMAP. In an embodiment, a variant of REMAP can be used to screen for compounds that bind to a variant of REMAP, but not to REMAP having the exact sequence of a sequence of SEQ ID NO:1-49. REMAP variants used to perform such screening can have a range of about 50% to about 99% sequence identity to REMAP, with various embodiments having 60%, 70%, 75%, 80%, 85%, 90%, and 95% sequence identity.

In an embodiment, a compound identified in a screen for specific binding to REMAP can be closely related to the natural ligand of REMAP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified can be a natural ligand of a receptor REMAP (Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246).

In other embodiments, a compound identified in a screen for specific binding to REMAP can

be closely related to the natural receptor to which REMAP binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for REMAP which is capable of propagating a signal, or a decoy receptor for REMAP which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Amgen Inc., Thousand Oaks CA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG_1 (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, two or more antibodies having similar or, alternatively, different specificities can be screened for specific binding to REMAP, fragments of REMAP, or variants of REMAP. The binding specificity of the antibodies thus screened can thereby be selected to identify particular fragments or variants of REMAP. In one embodiment, an antibody can be selected such that its binding specificity allows for preferential identification of specific fragments or variants of REMAP. In another embodiment, an antibody can be selected such that its binding specificity allows for preferential diagnosis of a specific disease or condition having increased, decreased, or otherwise abnormal production of REMAP.

In an embodiment, anticalins can be screened for specific binding to REMAP, fragments of REMAP, or variants of REMAP. Anticalins are ligand-binding proteins that have been constructed based on a lipocalin scaffold (Weiss, G.A. and H.B. Lowman (2000) Chem. Biol. 7:R177-R184; Skerra, A. (2001) J. Biotechnol. 74:257-275). The protein architecture of lipocalins can include a beta-barrel having eight antiparallel beta-strands, which supports four loops at its open end. These loops form the natural ligand-binding site of the lipocalins, a site which can be re-engineered *in vitro* by amino acid substitutions to impart novel binding specificities. The amino acid substitutions can be made using methods known in the art or described herein, and can include conservative substitutions (e.g., substitutions that do not alter binding specificity) or substitutions that modestly, moderately, or significantly alter binding specificity.

In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit REMAP involves producing appropriate cells which express REMAP, either as a secreted protein or on the cell membrane. Preferred cells can include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing REMAP or cell membrane fractions which contain REMAP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either REMAP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is

detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with REMAP, either in solution or affixed to a solid support, and detecting the binding of REMAP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

REMAP, fragments of REMAP, or variants of REMAP may be used to screen for compounds that modulate the activity of REMAP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for REMAP activity, wherein REMAP is combined with at least one test compound, and the activity of REMAP in the presence of a test compound is compared with the activity of REMAP in the absence of the test compound. A change in the activity of REMAP in the presence of the test compound is indicative of a compound that modulates the activity of REMAP. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising REMAP under conditions suitable for REMAP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of REMAP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding REMAP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (see, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R.

(1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding REMAP may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding REMAP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding REMAP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress REMAP, e.g., by secreting REMAP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of REMAP and receptors and membrane-associated proteins. In addition, examples of tissues expressing REMAP can be found in Table 6 and can also be found in Example XI. Therefore, REMAP appears to play a role in cell proliferative, autoimmune/inflammatory, neurological, metabolic, developmental, and endocrine disorders. In the treatment of disorders associated with increased REMAP expression or activity, it is desirable to decrease the expression or activity of REMAP. In the treatment of disorders associated with decreased REMAP expression or activity, it is desirable to increase the expression or activity of REMAP.

Therefore, in one embodiment, REMAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of REMAP. Examples of such disorders include, but are not limited to, a cell proliferative

disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia,

catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a metabolic disorder such as Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, lipid myopathies, lipodystrophies, lysosomal storage diseases, mannosidosis, neuraminidase deficiency, obesity, osteoporosis, phenylketonuria, pseudovitamin D-deficiency rickets, disorders of carbohydrate metabolism such as congenital type II dyserythropoietic anemia, diabetes, insulin-dependent diabetes mellitus, non-insulin-dependent diabetes mellitus, galactose epimerase deficiency, glycogen storage diseases, lysosomal storage diseases, fructosuria, pentosuria, and inherited abnormalities of pyruvate metabolism, disorders of lipid metabolism such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM_2 gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, and lipid myopathies, and disorders of copper metabolism such as Menke's disease, Wilson's disease, and Ehlers-Danlos syndrome type IX diabetes; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, a seizure disorder such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; and an endocrine disorder such as a disorder of the hypothalamus and/or pituitary resulting from lesions such as a primary brain tumor, adenoma, infarction associated with pregnancy, hypophysectomy, aneurysm, vascular malformation, thrombosis, infection, immunological disorder, and complication

due to head trauma, a disorder associated with hypopituitarism including hypogonadism, Sheehan syndrome, diabetes insipidus, Kallman's disease, Hand-Schuller-Christian disease, Letterer-Siwe disease, sarcoidosis, empty sella syndrome, and dwarfism, a disorder associated with hyperpituitarism including acromegaly, giantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma, a disorder associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection, subacute thyroiditis associated with viral infection, autoimmune thyroiditis (Hashimoto's disease), and cretinism, a disorder associated with hyperthyroidism including thyrotoxicosis and its various forms, Grave's disease, pretibial myxedema, toxic multinodular goiter, thyroid carcinoma, and Plummer's disease, a disorder associated with hyperparathyroidism including Conn disease (chronic hypercalemia), a pancreatic disorder such as Type I or Type II diabetes mellitus and associated complications, a disorder associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis, amyloidosis, hypokalemia, Cushing's disease, Liddle's syndrome, and Arnold-Healy-Gordon syndrome, pheochromocytoma tumors, and Addison's disease, a disorder associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, endometriosis, perturbation of the menstrual cycle, polycystic ovarian disease, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and, in post-menopausal women, osteoporosis, and, in men, Leydig cell deficiency, male climacteric phase, and germinal cell aplasia, a hypergonadal disorder associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5 α-reductase, and gynecomastia.

In another embodiment, a vector capable of expressing REMAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of REMAP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified REMAP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of REMAP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of REMAP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of REMAP including, but not limited to, those listed above.

In a further embodiment, an antagonist of REMAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of REMAP. Examples of such disorders include, but are not limited to, those cell proliferative, autoimmune/inflammatory, neurological, metabolic, developmental, and endocrine disorders described above. In one aspect, an

antibody which specifically binds REMAP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express REMAP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding REMAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of REMAP including, but not limited to, those described above.

In other embodiments, any protein, agonist, antagonist, antibody, complementary sequence, or vector embodiments may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of REMAP may be produced using methods which are generally known in the art. In particular, purified REMAP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind REMAP. Antibodies to REMAP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. In an embodiment, neutralizing antibodies (i.e., those which inhibit dimer formation) can be used therapeutically. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have application in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with REMAP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to REMAP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are substantially identical to a portion of the amino acid sequence of the natural protein.

Short stretches of REMAP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to REMAP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce REMAP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for REMAP may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 246:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between REMAP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering REMAP epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay

techniques may be used to assess the affinity of antibodies for REMAP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of REMAP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple REMAP epitopes, represents the average affinity, or avidity, of the antibodies for REMAP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular REMAP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the REMAP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of REMAP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of REMAP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty, *supra*; Coligan et al., *supra*).

In another embodiment of the invention, polynucleotides encoding REMAP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding REMAP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding REMAP (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102:469-475; Scanlon, K.J. et al. (1995) FASEB J. 9:1288-1296). Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) Blood 76:271-278;

Ausubel et al., *supra*; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63:323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (Rossi, J.J. (1995) Br. Med. Bull. 51:217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87:1308-1315; Morris, M.C. et al. (1997) Nucleic Acids Res. 25:2730-2736).

In another embodiment of the invention, polynucleotides encoding REMAP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in REMAP expression or regulation causes disease, the expression of REMAP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in REMAP are treated by constructing mammalian expression vectors encoding REMAP and introducing these vectors by mechanical means into REMAP-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J.-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of REMAP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (BD Clontech, Palo Alto CA).

REMAP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding REMAP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to REMAP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding REMAP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716;

Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In an embodiment, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding REMAP to cells which have one or more genetic abnormalities with respect to the expression of REMAP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999; Annu. Rev. Nutr. 19:511-544) and Verma, I.M. and N. Somia (1997; Nature 18:389:239-242).

In another embodiment, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding REMAP to target cells which have one or more genetic abnormalities with respect to the expression of REMAP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing REMAP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999; J. Virol. 73:519-532) and Xu, H. et al. (1994; Dev. Biol. 163:152-161). The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding REMAP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA,

resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for REMAP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of REMAP-coding RNAs and the synthesis of high levels of REMAP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of REMAP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of RNA molecules encoding REMAP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes may be prepared by any method

known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA molecules encoding REMAP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous endonucleases.

In other embodiments of the invention, the expression of one or more selected polynucleotides of the present invention can be altered, inhibited, decreased, or silenced using RNA interference (RNAi) or post-transcriptional gene silencing (PTGS) methods known in the art. RNAi is a post-transcriptional mode of gene silencing in which double-stranded RNA (dsRNA) introduced into a targeted cell specifically suppresses the expression of the homologous gene (i.e., the gene bearing the sequence complementary to the dsRNA). This effectively knocks out or substantially reduces the expression of the targeted gene. PTGS can also be accomplished by use of DNA or DNA fragments as well. RNAi methods are described by Fire, A. et al. (1998; Nature 391:806-811) and Gura, T. (2000; Nature 404:804-808). PTGS can also be initiated by introduction of a complementary segment of DNA into the selected tissue using gene delivery and/or viral vector delivery methods described herein or known in the art.

RNAi can be induced in mammalian cells by the use of small interfering RNA also known as siRNA. siRNA are shorter segments of dsRNA (typically about 21 to 23 nucleotides in length) that result *in vivo* from cleavage of introduced dsRNA by the action of an endogenous ribonuclease. siRNA appear to be the mediators of the RNAi effect in mammals. The most effective siRNAs appear to be 21 nucleotide dsRNAs with 2 nucleotide 3' overhangs. The use of siRNA for inducing RNAi in mammalian cells is described by Elbashir, S.M. et al. (2001; Nature 411:494-498).

siRNA can be generated indirectly by introduction of dsRNA into the targeted cell. Alternatively, siRNA can be synthesized directly and introduced into a cell by transfection methods and agents described herein or known in the art (such as liposome-mediated transfection, viral vector methods, or other polynucleotide delivery/introductory methods). Suitable siRNAs can be selected

by examining a transcript of the target polynucleotide (e.g., mRNA) for nucleotide sequences downstream from the AUG start codon and recording the occurrence of each nucleotide and the 3' adjacent 19 to 23 nucleotides as potential siRNA target sites, with sequences having a 21 nucleotide length being preferred. Regions to be avoided for target siRNA sites include the 5' and 3' untranslated regions (UTRs) and regions near the start codon (within 75 bases), as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP endonuclease complex. The selected target sites for siRNA can then be compared to the appropriate genome database (e.g., human, etc.) using BLAST or other sequence comparison algorithms known in the art. Target sequences with significant homology to other coding sequences can be eliminated from consideration. The selected siRNAs can be produced by chemical synthesis methods known in the art or by *in vitro* transcription using commercially available methods and kits such as the SILENCER siRNA construction kit (Ambion, Austin TX).

In alternative embodiments, long-term gene silencing and/or RNAi effects can be induced in selected tissue using expression vectors that continuously express siRNA. This can be accomplished using expression vectors that are engineered to express hairpin RNAs (shRNAs) using methods known in the art (see, e.g., Brummelkamp, T.R. et al. (2002) Science 296:550-553; and Paddison, P.J. et al. (2002) Genes Dev. 16:948-958). In these and related embodiments, shRNAs can be delivered to target cells using expression vectors known in the art. An example of a suitable expression vector for delivery of siRNA is the PSILENCER1.0-U6 (circular) plasmid (Ambion). Once delivered to the target tissue, shRNAs are processed *in vivo* into siRNA-like molecules capable of carrying out gene-specific silencing.

In various embodiments, the expression levels of genes targeted by RNAi or PTGS methods can be determined by assays for mRNA and/or protein analysis. Expression levels of the mRNA of a targeted gene can be determined, for example, by northern analysis methods using the NORTHERNMAX-GLY kit (Ambion); by microarray methods; by PCR methods; by real time PCR methods; and by other RNA/polynucleotide assays known in the art or described herein. Expression levels of the protein encoded by the targeted gene can be determined, for example, by microarray methods; by polyacrylamide gel electrophoresis; and by Western analysis using standard techniques known in the art.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding REMAP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide

sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased REMAP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding REMAP may be therapeutically useful, and in the treatment of disorders associated with decreased REMAP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding REMAP may be therapeutically useful.

In various embodiments, one or more test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding REMAP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding REMAP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding REMAP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved

using methods which are well known in the art (Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466).

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of REMAP, antibodies to REMAP, and mimetics, agonists, antagonists, or inhibitors of REMAP.

In various embodiments, the compositions described herein, such as pharmaceutical compositions, may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery allows administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising REMAP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, REMAP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell

culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example REMAP or fragments thereof, antibodies of REMAP, and agonists, antagonists or inhibitors of REMAP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind REMAP may be used for the diagnosis of disorders characterized by expression of REMAP, or in assays to monitor patients being treated with REMAP or agonists, antagonists, or inhibitors of REMAP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for REMAP include methods which utilize the antibody and a label to detect

REMAP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring REMAP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of REMAP expression. Normal or standard values for REMAP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to REMAP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of REMAP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, polynucleotides encoding REMAP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of REMAP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of REMAP, and to monitor regulation of REMAP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotides, including genomic sequences, encoding REMAP or closely related molecules may be used to identify nucleic acid sequences which encode REMAP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding REMAP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the REMAP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:50-98 or from genomic sequences including promoters, enhancers, and introns of the REMAP gene.

Means for producing specific hybridization probes for polynucleotides encoding REMAP include the cloning of polynucleotides encoding REMAP or REMAP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels,

such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding REMAP may be used for the diagnosis of disorders associated with expression of REMAP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders,

peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a metabolic disorder such as Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, lipid myopathies, lipodystrophies, lysosomal storage diseases, mannosidosis, neuraminidase deficiency, obesity, osteoporosis, phenylketonuria, pseudovitamin D-deficiency rickets, disorders of carbohydrate metabolism such as congenital type II dyserythropoietic anemia, diabetes, insulin-dependent diabetes mellitus, non-insulin-dependent diabetes mellitus, galactose epimerase deficiency, glycogen storage diseases, lysosomal storage diseases, fructosuria, pentosuria, and inherited abnormalities of pyruvate metabolism, disorders of lipid metabolism such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM_2 gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, and lipid myopathies, and disorders of copper metabolism such as Menke's disease, Wilson's disease, and Ehlers-Danlos syndrome type IX diabetes; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, a seizure disorder such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; and

an endocrine disorder such as a disorder of the hypothalamus and/or pituitary resulting from lesions such as a primary brain tumor, adenoma, infarction associated with pregnancy, hypophysectomy, aneurysm, vascular malformation, thrombosis, infection, immunological disorder, and complication due to head trauma, a disorder associated with hypopituitarism including hypogonadism, Sheehan syndrome, diabetes insipidus, Kallman's disease, Hand-Schuller-Christian disease, Letterer-Siwe disease, sarcoidosis, empty sella syndrome, and dwarfism, a disorder associated with hyperpituitarism including acromegaly, giantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma, a disorder associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection, subacute thyroiditis associated with viral infection, autoimmune thyroiditis (Hashimoto's disease), and cretinism, a disorder associated with hyperthyroidism including thyrotoxicosis and its various forms, Grave's disease, pretibial myxedema, toxic multinodular goiter, thyroid carcinoma, and Plummer's disease, a disorder associated with hyperparathyroidism including Conn disease (chronic hypercalemia), a pancreatic disorder such as Type I or Type II diabetes mellitus and associated complications, a disorder associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis, amyloidosis, hypokalemia, Cushing's disease, Liddle's syndrome, and Arnold-Healy-Gordon syndrome, pheochromocytoma tumors, and Addison's disease, a disorder associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, endometriosis, perturbation of the menstrual cycle, polycystic ovarian disease, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and, in post-menopausal women, osteoporosis, and, in men, Leydig cell deficiency, male climacteric phase, and germinal cell aplasia, a hypergonadal disorder associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5 α-reductase, and gynecomastia. Polynucleotides encoding REMAP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered REMAP expression. Such qualitative or quantitative methods are well known in the art.

In a particular embodiment, polynucleotides encoding REMAP may be used in assays that detect the presence of associated disorders, particularly those mentioned above. Polynucleotides complementary to sequences encoding REMAP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of polynucleotides encoding

REMAP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of REMAP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding REMAP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding REMAP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding REMAP, or a fragment of a polynucleotide complementary to the polynucleotide encoding REMAP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from polynucleotides encoding REMAP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers

derived from polynucleotides encoding REMAP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641).

Methods which may also be used to quantify the expression of REMAP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves (Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the

polynucleotides described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, REMAP, fragments of REMAP, or antibodies specific for REMAP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484; hereby expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson

(2000) Toxicol. Lett. 112-113:467-471). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity (see, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at niehs.nih.gov/oc/news/toxchip.htm). Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In an embodiment, the toxicity of a test compound can be assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another embodiment relates to the use of the polypeptides disclosed herein to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any

changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of interest. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for REMAP to quantify the levels of REMAP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by contacting the microarray with the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated

sample.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/25116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known and thoroughly described in Schena, M., ed. (1999; DNA Microarrays: A Practical Approach, Oxford University Press, London).

In another embodiment of the invention, nucleic acid sequences encoding REMAP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; Trask, B.J. (1991) Trends Genet. 7:149-154). Once mapped, the nucleic acid sequences may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP) (Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357).

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data (Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968). Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding REMAP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to

11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, REMAP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between REMAP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application WO84/03564). In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with REMAP, or fragments thereof, and washed. Bound REMAP is then detected by methods well known in the art. Purified REMAP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding REMAP specifically compete with a test compound for binding REMAP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with REMAP.

In additional embodiments, the nucleotide sequences which encode REMAP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/425,404, U.S. Ser. No. 60/440,907, U.S. Ser. No. 60/442,477, U.S. Ser. No. 60/448,565, U.S. Ser. No. 60/460,716, and U.S. Ser. No. 60/461,853, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs are derived from cDNA libraries described in the LIFESEQ database (Incyte, Palo Alto CA). Some tissues are homogenized and lysed in guanidinium isothiocyanate, while others are homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates are centrifuged over CsCl cushions or extracted with chloroform. RNA is precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA are repeated as necessary to increase RNA purity. In some cases, RNA is treated with DNase. For most libraries, poly(A)+ RNA is isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA is isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene is provided with RNA and constructs the corresponding cDNA libraries. Otherwise, cDNA is synthesized and cDNA libraries are constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription is initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters are ligated to double stranded cDNA, and the cDNA is digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA is size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs are ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen, Carlsbad CA), PCDNA2.1 plasmid (Invitrogen), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte, Palo Alto CA), pRARE (Incyte), or pINCY (Incyte), or derivatives thereof. Recombinant plasmids are transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Invitrogen.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I are recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids are purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids are resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA is amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps are carried out in a single reaction mixture. Samples are processed and stored in 384-well plates, and the concentration of amplified plasmid DNA is quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II are sequenced as follows. Sequencing reactions are processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions are prepared using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides are carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences are identified using standard methods (Ausubel et al., *supra*, ch. 7). Some of the cDNA sequences are selected for extension using the techniques disclosed in Example VIII.

Polynucleotide sequences derived from Incyte cDNAs are validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof are then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families; see, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries are performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences are assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs,

GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) are used to extend Incyte cDNA assemblages to full length. Assembly is performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages are screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences are translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences are subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (MiraiBio, Alameda CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences are also used to identify polynucleotide sequence fragments from SEQ ID NO:50-98. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative receptors and membrane-associated proteins are initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94; Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The

maximum range of sequence for Genscan to analyze at once is set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode receptors and membrane-associated proteins, the encoded polypeptides are analyzed by querying against PFAM models for receptors and membrane-associated proteins. Potential receptors and membrane-associated proteins are also identified by homology to Incyte cDNA sequences that have been annotated as receptors and membrane-associated proteins. These selected Genscan-predicted sequences are then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences are then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis is also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage is available, this information is used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences are obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences are derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences are extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III are mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster is analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that are subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval is present on more than one sequence in the cluster are identified, and intervals thus identified are considered to be equivalent by transitivity. For example, if an interval is present on a cDNA and two genomic sequences, then all three intervals are considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified are then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) are given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences are translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan are corrected by

comparison to the top BLAST hit from genpept. Sequences are further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences are extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III are queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog is then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein is generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both are used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences are therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences are examined to determine whether they contain a complete gene.

VI. Chromosomal Mapping of REMAP Encoding Polynucleotides

The sequences used to assemble SEQ ID NO:50-98 are compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:50-98 are assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon are used to determine if any of the clustered sequences have been previously mapped. Inclusion of a mapped sequence in a cluster results in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

Association of REMAP Polynucleotides with Parkinson's Disease

Several genes have been identified as showing linkage to autosomal dominant forms of Parkinson's Disease (PD). PD is a common neurodegenerative disorder causing bradykinesia, resting tremor, muscular rigidity, and postural instability. Cytoplasmic eosinophilic inclusions called Lewy bodies, and neuronal loss especially in the substantia nigra pars compacta, are pathological hallmarks of PD (Valente, E.M. et al (2001) Am. J. Hum. Genet. 68:895-900). Lewy body Parkinson disease has been thought to be a specific autosomal dominant disorder (Wakabayashi, K. et al. (1998) Acta Neuropath. 96:207-210). Juvenile parkinsonism may be a specific autosomal recessive disorder (Matsumine, H. et al. (1997) Am. J. Hum. Genet. 60: 588-596, 1997). (Online Mendelian Inheritance in Man, OMIM. Johns Hopkins University, Baltimore, MD. MIM Number: 168600: Sept. 9, 2002: . World Wide Web URL: http://www.ncbi.nlm.nih.gov/omim/)

Association of a disease with a chromosomal locus can be determined by lod score. Lod score is a statistical method used to test the linkage of two or more loci within families having a genetic disease. The lod score is the logarithm to base 10 of the odds in favor of linkage. Linkage is defined as the tendency of two genes located on the same chromosome to be inherited together through meiosis (*Genetics in Medicine*, Fifth Edition, (1991) Thompson, M.W. Et al. W.B. Saunders Co. Philadelphia). A lod score of +3 or greater (1000:1 odds in favor of linkage) indicates a probability of 1 in 1000 that a particular marker was found solely by chance in affected individuals, which is strong evidence that two genetic loci are linked.

One such gene implicated in PD is PARK3, which maps to 2p13 (Gasser, T. et al. (1998) Nature Genet. 18:262-265). A marker at chromosomal position D2S441 was found to have a lod score of 3.2 in the region of PARK3. This marker supported the disease association of PARK3 in the chromosomal interval from D2S134 to D2S286 (Gasser et al., supra). Markers located within chromosomal intervals D2S134 and D2S286, which map between 83.88 to 94.05 centiMorgans on the short arm of chromosome 2, were used to identify genes that map in the region between D2S134 and D2S286.

A second PD gene, implicated in early-onset recessive parkinsonism, is PARK6, located on chromosome 1 at 1p35-1p36. Several markers were obtained with lod scores greater than 3 including D1S199, D1S2732, D1S2828, D1S478, D1S2702, D1S2734, D1S2674 (Valente, E.M. et al. supra). These markers were used to determine the PD-relevant range of chromosome loci and identify sequences that map to chromosome 1 between D1S199 and D1S2885. REMAP polynucleotides were found to map within the chromosomal region in which markers associated with disease or other physiological processes of interest were located. Genomic contigs available from NCBI were used to identify REMAP polynucleotides which map to a disease locus. Contigs longer than 1Mb were broken into subcontigs of 1Mb in length with overlapping sections of 100 kb. A preliminary step

used an algorithm, similar to MEGABLAST (NCBI), to identify mRNA sequence/masked genomic DNA contig pairings. SIM4 (Florea, L. et al. (1998) Genome Res. 8:967-74, version May 2000) was optimized for high throughput and strand assignment confidence, and used to further select cDNA/genomic pairings. The SIM4-selected mRNA sequence/genomic contig pairs were further processed to determine the correct location of the REMAP polynucleotides on the genomic contig and their strand identity.

SEQ ID NO:50 mapped to a region of contig NT_030563.2 from the February 2002 GenBank release, localizing SEQ ID NO:50 to within 14.9 MB of the Parkinson's disease locus on chromosome 1, a chromosomal region consistently associated with Parkinson's disease.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook and Russell, *supra*, ch. 7; Ausubel et al.; *supra*, ch. 4).

Analogous computer techniques applying BLAST are used to search for identical or related molecules in databases such as GenBank or LIFESEQ (Incyte). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79%

identity and 100% overlap.

Alternatively, polynucleotides encoding REMAP are analyzed with respect to the tissue sources from which they are derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding REMAP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ database (Incyte, Palo Alto CA).

VIII. Extension of REMAP Encoding Polynucleotides

Full length polynucleotides are produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer is synthesized to initiate 5' extension of the known fragment, and the other primer is synthesized to initiate 3' extension of the known fragment. The initial primers are designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations is avoided.

Selected human cDNA libraries are used to extend the sequence. If more than one extension is necessary or desired, additional or nested sets of primers are designed.

High fidelity amplification is obtained by PCR using methods well known in the art. PCR is performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contains DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ are as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec;

Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well is determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate is scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture is analyzed by electrophoresis on a 1 % agarose gel to determine which reactions are successful in extending the sequence.

The extended nucleotides are desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun sequencing, the digested nucleotides are separated on low concentration (0.6 to 0.8%) agarose gels, fragments are excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells are selected on antibiotic-containing media, and individual colonies are picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells are lysed, and DNA is amplified by PCR using Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA is quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries are reamplified using the same conditions as described above. Samples are diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotides are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in REMAP Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) are identified in SEQ ID NO:50-98 using the LIFESEQ database (Incyte). Sequences from the same

gene are clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters is used to distinguish SNPs from other sequence variants. Preliminary filters remove the majority of basecall errors by requiring a minimum Phred quality score of 15, and remove sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis is applied to the original chromatogram files in the vicinity of the putative SNP. Clone error filters use statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters use statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removes duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs are selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprises 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprises 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprises 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprises 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies are first analyzed in the Caucasian population; in some cases those SNPs which show no allelic variance in this population are not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:50-98 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to NYTRAN

PLUS nylon membranes (Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler et al., *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena, M., ed. (1999) DNA Microarrays: A Practical Approach, Oxford University Press, London). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31).

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse

transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (BD Clontech, Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma-Aldrich, St. Louis MO) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered

with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and

measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte). Array elements that exhibit at least about a two-fold change in expression, a signal-to-background ratio of at least about 2.5, and an element spot size of at least about 40%, are considered to be differentially expressed.

Expression

In an example, SEQ ID NO:64 showed differential expression in lung cancer samples, as determined by microarray analysis. Grossly uninvolved lung tissue from a number of different donors was compared to lung tumor tissue derived from the same donors (Roy Castle International Centre for Lung Cancer Research, Liverpool, UK). Expression of SEQ ID NO:64 was shown to be decreased by at least 2-fold in tumor tissue when compared to the matched normal tissue from the same donor. This decreased expression was found in 2 out of 10 total matched donor samples. Therefore, in various embodiments, SEQ ID NO:64 can be used for one or more of the following: i) monitoring treatment of lung cancer, ii) diagnostic assays for lung cancer, and iii) developing therapeutics and/or other treatments for lung cancer.

In another example, SEQ ID NO:64 showed differential expression in ovarian cancer tissue samples, as determined by microarray analysis. A normal ovary from a 79 year-old female donor was compared to an ovarian tumor from the same donor (Huntsman Cancer Institute, Salt Lake City, UT). The expression of SEQ ID NO:64 was decreased at least 3.5-fold in the tumor tissue, when compared to tissue from the normal ovary. Therefore, in various embodiments, SEQ ID NO:64 can be used for one or more of the following: i) monitoring treatment of ovarian cancer, ii) diagnostic assays for ovarian cancer, and iii) developing therapeutics and/or other treatments for ovarian cancer.

In another example, SEQ ID NO:66 was differentially expressed in several breast cancer cell lines, as determined by microarray analysis. The gene expression profile of a nonmalignant mammary epithelial cell line was compared to the gene expression profiles of breast carcinoma lines at different stages of tumor progression. Cell lines compared included: a) MCF-10A, a breast mammary gland (luminal ductal characteristics) cell line isolated from a 36-year-old woman with fibrocystic breast disease, b) MCF7, a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year-old female, c) BT-20, a breast carcinoma cell line derived *in vitro* from the cells emigrating out of thin slices of tumor mass isolated from a 74-year-old female, d) BT-474, a breast ductal carcinoma cell line that was isolated from a solid, invasive ductal carcinoma of the

breast obtained from a 60-year-old woman, e) BT-483, a breast ductal carcinoma cell line that was isolated from a papillary invasive ductal tumor obtained from a 23-year-old normal, menstruating, parous female with a family history of breast cancer, f) Hs 578T, a breast ductal carcinoma cell line isolated from a 74-year-old female with breast carcinoma, and g) MDA-MB-468, a breast adenocarcinoma cell line isolated from the pleural effusion of a 51-year-old female with metastatic adenocarcinoma of the breast. All of these cell lines were compared with HMEC, a primary breast epithelial cell line isolated from a normal donor. In one set of experiments, all cells were grown under optimal growth conditions, in the presence of growth factors and nutrients. The expression levels of SEQ ID NO:66 were decreased at least 2.5-fold in Hs 578T and MDA-MB-468 cells, at least 3-fold in MCF-10A and MCF7 cells, at least 4-fold in BT-20 cells, and at least 5-fold in BT-474 and BT-483 cells, when compared with expression levels in HMEC cells. In another set of experiments, all cells were grown in basal media in the absence of growth factors and hormones for 24 hours prior to comparison. The expression levels of SEQ ID NO:66 were decreased at least 2-fold in Hs 578T cells, at least 3-fold in BT-474 cells, at least 4-fold in MDA-MB-468, BT-20 and BT-483 cells, and at least 5-fold in MCF7 cells, when compared to expression levels in HMEC cells. Therefore, in various embodiments, SEQ ID NO:66 can be used for one or more of the following: i) monitoring treatment of breast cancer, ii) diagnostic assays for breast cancer, and iii) developing therapeutics and/or other treatments for breast cancer.

In another example, SEQ ID NO:66 showed differential expression in colon cancer tissue samples, as determined by microarray analysis. Gene expression profiles were obtained by comparing normal colon tissue from several donors to colon tumor tissue isolated from the same donors (Huntsman Cancer Institute, Salt Lake City, UT). In 2 out of 5 colon tumor samples, the expression of SEQ ID NO:66 was increased at least 2-fold when compared to normal colon tissue isolated from the same donor. Therefore, in various embodiments, SEQ ID NO:66 can be used for one or more of the following: i) monitoring treatment of colon cancer, ii) diagnostic assays for colon cancer, and iii) developing therapeutics and/or other treatments for colon cancer.

In another example, SEQ ID NO:66 showed differential expression in several bone tumor samples, as determined by microarray expression analysis. Messenger RNA from normal human osteoblasts was compared with mRNA from biopsy specimens, osteosarcoma tissues, or primary cultures or metastasized tissues. A normal osteoblast primary culture, NHOst 5488, was chosen as the reference in the initial experiments. One basic set of experiments is defined as the comparison of mRNA from biopsy specimen with that of definitive surgical specimen from the same patient. Extended study of this basic set includes mRNA from primary cell cultures of the definitive surgical specimen, muscle, or cartilage tissue from the same patient. Biopsy specimens, definitive surgical specimens, or lung metastatic tissues from different individuals were also included to reveal

individual variability. In 5 different donors, the average expression level of SEQ ID NO:66 in the tumor samples from each donor was increased at least 2-fold in comparison with the gene expression levels detected in normal human osteoblasts. Therefore, in various embodiments, SEQ ID NO:66 can be used for one or more of the following: i) monitoring treatment of osteosarcoma, ii) diagnostic assays for osteosarcoma, and iii) developing therapeutics and/or other treatments for osteosarcoma.

In another example, SEQ ID NO:67 showed tissue-specific expression as determined by microarray analysis. RNA samples isolated from a variety of normal human tissues were compared to a common reference sample. Tissues contributing to the reference sample were selected for their ability to provide a complete distribution of RNA in the human body and include brain (4%), heart (7%), kidney (3%), lung (8%), placenta (46%), small intestine (9%), spleen (3%), stomach (6%), testis (9%), and uterus (5%). The normal tissues assayed were obtained from at least three different donors. RNA from each donor was separately isolated and individually hybridized to the microarray. Since these hybridization experiments were conducted using a common reference sample, differential expression values are directly comparable from one tissue to another. The expression of SEQ ID NO:67 was increased by at least 2-fold in brain tissue samples as compared to the reference sample. Therefore, SEQ ID NO:67 can be used as a tissue marker for brain. The expression of SEQ ID NO:67 was increased by at least 2-fold in thymus glad as compared to the reference sample. Therefore, SEQ ID NO:67 can be used as a tissue marker for thymus gland.

In another example, SEQ ID NO:75 showed tissue-specific expression as determined by microarray analysis. RNA samples isolated from a variety of normal human tissues were compared to a common reference sample. Tissues contributing to the reference sample were selected for their ability to provide a complete distribution of RNA in the human body and include brain (4%), heart (7%), kidney (3%), lung (8%), placenta (46%), small intestine (9%), spleen (3%), stomach (6%), testis (9%), and uterus (5%). The normal tissues assayed were obtained from at least three different donors. RNA from each donor was separately isolated and individually hybridized to the microarray. Since these hybridization experiments were conducted using a common reference sample, differential expression values are directly comparable from one tissue to another. The expression of SEQ ID NO:75 was increased by at least two-fold in ileal tissue as compared to the reference sample. Therefore, SEQ ID NO:75 can be used as a tissue marker for ileal tissue.

As another example, SEQ ID NO:71 showed tissue-specific expression as determined by microarray analysis. Specific dissected brain regions from a 55-year-old non-demented human female brain form the basis of this anatomical survey. Brain regions were pooled for use as the control (Loyola Brain Bank, Chicago IL). The mixed brain control included: Amygdala, Entorhinal Cortex (8.33%), Globus Pallidus & Substantia Innominata (4.17%), Striatum, Dorsal Caudate Nucleus, Dorsal Putamen, Ventral Nucleus Accumbens (4.17%), Archaecortex, Hippocampus

Anterior (4.17%), Archaecortex, Hippocampus Posterior (4.17%), Neocortex, Auditory (8.33%), Neocortex, Frontal (8.33%), Neocortex, Sensory, Motor (8.33%), Neocortex, Temporal (8.33%), Neocortex, Visual Primary (8.33%), Thalamus (8.33%), Vermis (8.33%), Pons (Brainstem) (8.33%), Medulla (Brainstem) (8.33%). The expression of SEQ ID NO:71 was decreased by at least two-fold in vermis tissue and dentate nucleus tissue as compared to the reference sample. Therefore, SEQ ID NO:71 can be used for one or more of the following: i) monitoring treatment of neurological disorders and related diseases and conditions, ii) diagnostic assays for neurological disorders and related diseases and conditions, and iii) developing therapeutics and/or other treatments for neurological disorders and related diseases and conditions.

As another example, expression of SEQ ID NO:71 was upregulated in peripheral blood mononuclear cells (PBMCs) treated with cytokines versus untreated PBMCs as determined by microarray analysis. In one experiment, PBMCs collected from the blood of 6 healthy donors were placed in culture for 2 hours in the presence or absence of 10 ng/ml recombinant interleukin 1 beta (IL-1β). IL-1β treated PBMCs and untreated control PBMCs from the different donors were pooled according to their respective treatment. Expression of SEQ ID NO:71 was increased at least two-fold in 1 of 1 samples tested. In one experiment, PBMCs collected from the blood of 6 healthy donors were placed in culture for 2 hours in the presence or absence of 10 ng/ml recombinant interleukin 6 (IL-6). IL-6 treated PBMCs and untreated control PBMCs from the different donors were pooled according to their respective treatment. Expression of SEQ ID NO:71 was increased at least two-fold in 1 of 1 samples tested. In a second experiment, PBMCs were collected from the blood of 6 healthy volunteer donors using standard gradient separation. The PBMCs from each donor were placed in culture for 2 hours in the presence or absence of 10 ng/ml recombinant interleukin 5 (IL-5). IL-5 treated PBMCs and untreated control PBMCs from the different donors were pooled according to their respective treatments. Expression of SEQ ID NO:71 was increased at least two-fold in 1 of 2 samples tested. Therefore, in various embodiments, SEQ ID NO:71 can be used for one or more of the following: i) monitoring treatment of immune disorders and related diseases and conditions, ii) diagnostic assays for immune disorders and related diseases and conditions, and iii) developing therapeutics and/or other treatments for immune disorders and related diseases and conditions.

As another example, expression of SEQ ID NO:71 was upregulated in breast ductal carcinoma cells versus nonmalignant mammary epithelial cells as determined by microarray analysis. The gene expression profile of a nonmalignant mammary epithelial cell line was compared to the gene expression profiles of breast carcinoma lines at different stages of tumor progression. Cell lines compared included: a) MCF-10A, a breast mammary gland (luminal ductal characteristics) cell line isolated from a 36-year-old woman with fibrocystic breast disease, b)MCF7, a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year-old female, c) BT-20, a

breast carcinoma cell line derived *in vitro* from the cells emigrating out of thin slices of tumor mass isolated from a 74-year-old female, d) BT-474, a breast ductal carcinoma cell line that was isolated from a solid, invasive ductal carcinoma of the breast obtained from a 60-year-old woman, e) BT-483, a breast ductal carcinoma cell line that was isolated from a papillary invasive ductal tumor obtained from a 23-year-old normal, menstruating, parous female with a family history of breast cancer, f) Hs 578T, a breast ductal carcinoma cell line isolated from a 74-year-old female with breast carcinoma, and g) MDA-MB-468, a breast adenocarcinoma cell line isolated from the pleural effusion of a 51-year-old female with metastatic adenocarcinoma of the breast. In one experiment, all cells were grown under optimal growth conditions, in the presence of growth factors and nutrients. In a second experiment, all cells were grown in basal media in the absence of growth factors and hormones for 24 hours prior to comparison. Under optimal growth conditions, expression of SEQ ID NO:71 was increased at least two-fold in 1 of 2 Hs 578T samples tested. In the absence of growth factors and hormones, expression of SEQ ID NO:71 was increased at least two-fold in 1 of 2 Hs 578T samples tested. Therefore, in various embodiments, SEQ ID NO:71 can be used for one or more of the following: i) monitoring treatment of breast cancer, ii) diagnostic assays for breast cancer, and iii) developing therapeutics and/or other treatments for breast cancer.

As another example, expression of SEQ ID NO:71 was upregulated in prostate carcinoma cells versus nonmalignant mammary epithelial cells as determined by microarray analysis. Primary prostate epithelial cells were compared with prostate carcinomas representative of the different stages of tumor progression. Cell lines compared included: a) PrEC, a primary prostate epithelial cell line isolated from a normal donor, b) DU 145, a prostate carcinoma cell line isolated from a metastatic site in the brain of 69-year old male with widespread metastatic prostate carcinoma, c) LNCaP, a prostate carcinoma cell line isolated from a lymph node biopsy of a 50-year-old male with metastatic prostate carcinoma, d) PC-3, a prostate adenocarcinoma cell line isolated from a metastatic site in the bone of a 62-year-old male with grade IV prostate adenocarcinoma, and e) MDAPCa2b, a prostate adenocarcinoma cell line isolated from a metastatic site in the bone of a 63-year-old male. Cells grown under restrictive conditions were compared to normal PrECs grown under restrictive conditions. In one experiment, cells were grown in basal media in the absence of growth factors and hormones. In another experiment, cells were grown under optimal growth conditions, in the presence of growth factors and nutrients. Under optimal growth conditions, expression of SEQ ID NO:71 was increased at least two-fold in 2 of 2 DU 145 samples tested. In the absence of growth factors and hormones, expression of SEQ ID NO:71 was increased at least two-fold in 1 of 1 DU 145 samples tested. Therefore, in various embodiments, SEQ ID NO:71 can be used for one or more of the following: i) monitoring treatment of prostate cancer, ii) diagnostic assays for prostate cancer, and iii) developing therapeutics and/or other treatments for prostate cancer.

In another example, SEQ ID NO:84 showed differential expression in activated immune system cells as determined by microarray analysis. In one experiment, SEQ ID NO:84 showed differential expression in activated T cells. PHA blasts derived from the PBMCs of 5 healthy volunteer donors were stimulated for 9 days in the presence of PHA and IL-2. These T cell blasts were washed and stimulated for 1 and 6 hours in the presence of anti-CD3 monoclonal antibody, anti-CD28 antibody, or a combination of both. These reactivated T cells were compared to matching untreated PHA blasts. Treatment of the T cell blasts with a combination of the two antibodies for 6 hours led to an increase in SEQ ID NO:84 expression of at least 2-fold, when compared to untreated PHA blasts. This increase in SEQ ID NO:84 expression did not occur when T cell blasts were treated with each antibody alone. In another experiment, SEQ ID NO:84 showed differential expression in activated PBMCs. Human peripheral blood mononuclear cells (PBMCs) contain about 52% lymphocytes (12% B lymphocytes, 40% T lymphocytes {25% CD4+ and 15% CD8+}), 20% NK cells, 25% monocytes, and 3% various cells that include dendritic cells and progenitor cells. Treatment of PBMCs with a combination of PMA and ionomycin mimics the secondary signaling events elicited during activation of lymphocytes, NK cells, and monocytes. PBMCs were stimulated *in vitro* with 0.1 μM/ml soluble phorbol myristate acetate (PMA) and 10 ng/ml ionomycin for 2, 4, 8, and 20 hours. Treated cells were compared to untreated PBMCs kept in culture in the absence of stimuli. Expression of SEQ ID NO:84 increased at least 2-fold after 4 hours of treatment with PMA and ionomycin, and continued to increase to at least 4-fold after 20 hours of treatment, when compared to SEQ ID NO:84 expression levels in untreated PBMCs. In another experiment, SEQ ID NO:84 showed differential expression in PBMCs treated with SEB. Peripheral blood mononuclear cells (PBMCs) from healthy volunteer donors were activated with 1 ng/ml Staphylococcal exotoxin B (SEB) for 72 hours. The treated PBMCs were compared to matching PBMCs kept in culture in the presence of medium alone. Expression of SEQ ID NO:84 increased at least 2-fold in the SEB-treated PBMCs, when compared to the expression levels detected in the untreated cells. Therefore, in various embodiments, SEQ ID NO:84 can be used for one or more of the following: i) monitoring treatment of immune disorders and related diseases and conditions, ii) diagnostic assays for immune disorders and related diseases and conditions, and iii) developing therapeutics and/or other treatments for immune disorders and related diseases and conditions.

In another example, SEQ ID NO:89 showed tissue-specific expression as determined by microarray analysis. RNA samples isolated from a variety of normal human tissues were compared to a common reference sample. Tissues contributing to the reference sample were selected for their ability to provide a complete distribution of RNA in the human body and include brain (4%), heart (7%), kidney (3%), lung (8%), placenta (46%), small intestine (9%), spleen (3%), stomach (6%), testis (9%), and uterus (5%). The normal tissues assayed were obtained from at least three different

donors. RNA from each donor was separately isolated and individually hybridized to the microarray. Since these hybridization experiments were conducted using a common reference sample, differential expression values are directly comparable from one tissue to another. The expression of SEQ ID NO:89 was increased by at least two-fold in brain and thymus gland tissues as compared to the reference sample. Therefore, SEQ ID NO:89 can be used as a tissue marker for brain and thymus gland tissues.

In another example, SEQ ID NO:92 showed differential expression in prostate cancer cell lines, as determined by microarray expression analysis. Primary prostate epithelial cells (PrEC) were compared with prostate carcinomas representative of the different stages of tumor progression. Cell lines compared included: a) DU 145, a prostate carcinoma cell line isolated from a metastatic site in the brain of 69-year old male with widespread metastatic prostate carcinoma, b) LNCaP, a prostate carcinoma cell line isolated from a lymph node biopsy of a 50-year-old male with metastatic prostate carcinoma, and c) PC-3, a prostate adenocarcinoma cell line isolated from a metastatic site in the bone of a 62-year-old male with grade IV prostate adenocarcinoma. In one experiment, all cell lines were grown in basal media in the absence of growth factors and hormones. Expression of SEQ ID NO:92 was decreased at least 2.5-fold in the prostate cancer cell lines PC3, LNCaP, and DU 145, when compared to expression levels detected in the control PrEC cell line. In another experiment, the prostate cancer cell lines were grown under optimal growth conditions, in the presence of growth factors and nutrients, and compared to starved PrEC cells. The expression of SEQ ID NO:92 was decreased at least 2.5-fold in DU 145 cells and at least 3-fold in the PC3 and LNCaP cell lines, when compared to the expression levels of SEQ ID NO:92 detected in the PrEC cells. Therefore, in various embodiments, SEQ ID NO:92 can be used for one or more of the following: i) monitoring treatment of prostate cancer, ii) diagnostic assays for prostate cancer, and iii) developing therapeutics and/or other treatments for prostate cancer.

In another example, expression of SEQ ID NO:95 was downregulated in treated PBMC cells versus untreated PBMC cells as determined by microarray analysis. PBMCs from the blood of six healthy volunteer donors were incubated for 24 hours in the presence of graded doses of beclomethasone dissolved in DMSO. In addition, matching PBMCs were treated for the same duration with matching doses of DMSO in order to monitor the possible effects of the vehicle alone. The treated PBMC were compared to matching untreated PBMC maintained in culture for the same duration. Expression of SEQ ID NO:95 was decreased at least two-fold in pooled PBMC cells treated with 0.2 μM, 1 μM, and 5 μM beclomethasone when compared to untreated pooled PBMC cells. Therefore, in various embodiments, SEQ ID NO:95 can be used for one or more of the following: i) monitoring treatment of immune disorders and related diseases and conditions, ii) diagnostic assays for immune disorders and related diseases and conditions, and iii) developing therapeutics and/or

other treatments for immune disorders and related diseases and conditions.

In another example, expression of SEQ ID NO:95 was upregulated in cancerous colon tissue versus normal colon tissue as determined by microarray analysis. Matched normal and tumor samples from a 58-year-old female diagnosed with mucinous adenocarcinoma (Huntsman Cancer Institute, Salt Lake City, UT) were compared by competitive hybridization. Expression of SEQ ID NO:95 was increased at least two-fold in colon adenocarcinoma tissue when compared to grossly uninvolved collon tissue from the same donor. Therefore, in various embodiments, SEQ ID NO:95 can be used for one or more of the following: i) monitoring treatment of colon cancer, ii) diagnostic assays for colon cancer, and iii) developing therapeutics and/or other treatments for colon cancer.

In yet another example, SEQ ID NO:94 and SEQ ID NO:98 showed tissue-specific expression as determined by microarray analysis. RNA samples isolated from a variety of normal human tissues were compared to a common reference sample. Tissues contributing to the reference sample were selected for their ability to provide a complete distribution of RNA in the human body and include brain (4%), heart (7%), kidney (3%), lung (8%), placenta (46%), small intestine (9%), spleen (3%), stomach (6%), testis (9%), and uterus (5%). The normal tissues assayed were obtained from at least three different donors. RNA from each donor was separately isolated and individually hybridized to the microarray. Since these hybridization experiments were conducted using a common reference sample, differential expression values are directly comparable from one tissue to another. The expression of SEQ ID NO:94 was increased by at least two-fold in liver as compared to the reference sample. Further, the expression of SEQ ID NO:98 was increased by at least two-fold in brain as compared to the reference sample. Therefore, SEQ ID NO:94 can be used as a tissue marker for liver and SEQ ID NO:98 can be used as a tissue marker for brain.

XII. Complementary Polynucleotides

Sequences complementary to the REMAP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring REMAP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of REMAP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the REMAP-encoding transcript.

XIII. Expression of REMAP

Expression and purification of REMAP is achieved using bacterial or virus-based expression systems. For expression of REMAP in bacteria, cDNA is subcloned into an appropriate vector

containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express REMAP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of REMAP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding REMAP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus (Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945).

In most expression systems, REMAP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences). Following purification, the GST moiety can be proteolytically cleaved from REMAP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). Purified REMAP obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, and XIX, where applicable.

XIV. Functional Assays

REMAP function is assessed by expressing the sequences encoding REMAP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an

additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; BD Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994; Flow Cytometry, Oxford, New York NY).

The influence of REMAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding REMAP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding REMAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of REMAP Specific Antibodies

REMAP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the REMAP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art (Ausubel et al., *supra*, ch. 11).

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to

increase immunogenicity (Ausubel et al., *supra*). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-REMAP activity by, for example, binding the peptide or REMAP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring REMAP Using Specific Antibodies

Naturally occurring or recombinant REMAP is substantially purified by immunoaffinity chromatography using antibodies specific for REMAP. An immunoaffinity column is constructed by covalently coupling anti-REMAP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing REMAP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of REMAP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/REMAP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and REMAP is collected.

XVII. Identification of Molecules Which Interact with REMAP

REMAP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled REMAP, washed, and any wells with labeled REMAP complex are assayed. Data obtained using different concentrations of REMAP are used to calculate values for the number, affinity, and association of REMAP with the candidate molecules.

Alternatively, molecules interacting with REMAP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989; Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (BD Clontech).

REMAP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of REMAP Activity

An assay for REMAP activity measures the expression of REMAP on the cell surface. cDNA encoding REMAP is transfected into an appropriate mammalian cell line. Cell surface proteins are labeled with biotin as described (de la Fuente, M.A. et al. (1997) Blood 90:2398-2405). Immunoprecipitations are performed using REMAP-specific antibodies, and immunoprecipitated samples are analyzed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled immunoprecipitant is proportional to the amount of REMAP expressed on the cell surface.

In the alternative, an assay for REMAP activity is based on a prototypical assay for ligand/receptor-mediated modulation of cell proliferation. This assay measures the rate of DNA synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding REMAP is added to quiescent 3T3 cultured cells using transfection methods well known in the art. The transiently transfected cells are then incubated in the presence of [^{3}H]thymidine, a radioactive DNA precursor molecule. Varying amounts of REMAP ligand are then added to the cultured cells. Incorporation of [^{3}H]thymidine into acid-precipitable DNA is measured over an appropriate time interval using a radioisotope counter, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold REMAP ligand concentration range is indicative of receptor activity. One unit of activity per milliliter is defined as the concentration of REMAP producing a 50% response level, where 100% represents maximal incorporation of [^{3}H]thymidine into acid-precipitable DNA (McKay, I. and I. Leigh, eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York NY, p. 73.)

In a further alternative, the assay for REMAP activity is based upon the ability of GPCR family proteins to modulate G protein-activated second messenger signal transduction pathways (e.g., cAMP; Gaudin, P. et al. (1998) J. Biol. Chem. 273:4990-4996). A plasmid encoding full length REMAP is transfected into a mammalian cell line (e.g., Chinese hamster ovary (CHO) or human embryonic kidney (HEK-293) cell lines) using methods well-known in the art. Transfected cells are grown in 12-well trays in culture medium for 48 hours, then the culture medium is discarded, and the attached cells are gently washed with PBS. The cells are then incubated in culture medium with or without ligand for 30 minutes, then the medium is removed and cells lysed by treatment with 1 M perchloric acid. The cAMP levels in the lysate are measured by radioimmunoassay using methods well-known in the art. Changes in the levels of cAMP in the lysate from cells exposed to ligand compared to those without ligand are proportional to the amount of REMAP present in the transfected cells.

To measure changes in inositol phosphate levels, the cells are grown in 24-well plates containing $1x10^5$ cells/well and incubated with inositol-free media and [^{3}H]myoinositol, 2 μCi/well, for 48 hr. The culture medium is removed, and the cells washed with buffer containing 10 mM LiCl followed by addition of ligand. The reaction is stopped by addition of perchloric acid. Inositol phosphates are extracted and separated on Dowex AG1-X8 (Bio-Rad) anion exchange resin, and the total labeled inositol phosphates counted by liquid scintillation. Changes in the levels of labeled inositol phosphate from cells exposed to ligand compared to those without ligand are proportional to

the amount of REMAP present in the transfected cells.

In a further alternative, the ion conductance capacity of REMAP is demonstrated using an electrophysiological assay. REMAP is expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector encoding REMAP. Eukaryotic expression vectors are commercially available, and the techniques to introduce them into cells are well known to those skilled in the art. A small amount of a second plasmid, which expresses any one of a number of marker genes such as β-galactosidase, is co-transformed into the cells in order to allow rapid identification of those cells which have taken up and expressed the foreign DNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow expression and accumulation of REMAP and β-galactosidase. Transformed cells expressing β-galactosidase are stained blue when a suitable colorimetric substrate is added to the culture media under conditions that are well known in the art. Stained cells are tested for differences in membrane conductance due to various ions by electrophysiological techniques that are well known in the art. Untransformed cells, and/or cells transformed with either vector sequences alone or β-galactosidase sequences alone, are used as controls and tested in parallel. The contribution of REMAP to cation or anion conductance can be shown by incubating the cells using antibodies specific for either REMAP. The respective antibodies will bind to the extracellular side of REMAP, thereby blocking the pore in the ion channel, and the associated conductance.

In a further alternative, REMAP transport activity is assayed by measuring uptake of labeled substrates into *Xenopus laevis* oocytes. Oocytes at stages V and VI are injected with REMAP mRNA (10 ng per oocyte) and incubated for 3 days at 18 °C in OR2 medium (82.5 mM NaCl, 2.5 mM KCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 1 mM Na_2HPO_4, 5 mM Hepes, 3.8 mM NaOH , 50 μg/ml gentamycin, pH 7.8) to allow expression of REMAP protein. Oocytes are then transferred to standard uptake medium (100 mM NaCl, 2 mM KCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 10 mM Hepes/Tris pH 7.5). Uptake of various substrates (e.g., amino acids, sugars, drugs, and neurotransmitters) is initiated by adding a ^3H substrate to the oocytes. After incubating for 30 minutes, uptake is terminated by washing the oocytes three times in Na^+-free medium, measuring the incorporated ^3H, and comparing with controls. REMAP activity is proportional to the level of internalized ^3H substrate.

In a further alternative, REMAP protein kinase (PK) activity is measured by phosphorylation of a protein substrate using gamma-labeled [^{32}P]-ATP and quantitation of the incorporated radioactivity using a gamma radioisotope counter. REMAP is incubated with the protein substrate, [^{32}P]-ATP, and an appropriate kinase buffer. The ^{32}P incorporated into the product is separated from free [^{32}P]-ATP by electrophoresis and the incorporated ^{32}P is counted. The amount of ^{32}P recovered is proportional to the PK activity of REMAP in the assay. A determination of the specific amino acid residue phosphorylated is made by phosphoamino acid analysis of the hydrolyzed protein.

In another alternative, REMAP activity is measured by its ability to stimulate transcription of a reporter gene in respone to hormone ligands. The assay entails the use of a well characterized reporter gene construct containing the luciferase gene under control of a basal transcriptional promoter. The methods for constructing and expressing fusion genes, introducing them into cells, and measuring luciferase enzyme activity, are well known to those skilled in the art. In preparation for this assay, REMAP is expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector containing REMAP cDNA, along with the reporter gene construct. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow expression and accumulation of REMAP. The cells are then treated with REMAP hormone ligand, or left untreated. Cell lysates are then prepared according to the luciferase assay method to be used, several of which are well known to those skilled in the art. Comparison of the hormone ligand-stimulated luciferase activity with the activity seen in unstimulated cells provides a measure of REMAP activity.

In the alternative, a method to determine nucleic acid binding activity of REMAP in response to hormone ligands involves a polyacrylamide gel mobility-shift assay. In preparation for this assay, REMAP is expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector containing REMAP cDNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow expression and accumulation of REMAP. Cells are then treated with media containing REMAP hormone ligand, or with media alone. Extracts containing solubilized proteins can be prepared from cells expressing REMAP by methods well known in the art. Portions of the extract containing REMAP are added to [^{32}P]-labeled RNA or DNA. Radioactive nucleic acid can be synthesized *in vitro* by techniques well known in the art. The mixtures are incubated at 25°C in the presence of RNase- and DNase-inhibitors under buffered conditions for 5-10 minutes. After incubation, the samples are analyzed by polyacrylamide gel electrophoresis followed by autoradiography. The presence of a band on the autoradiogram indicates the formation of a complex between REMAP and the radioactive transcript. A band of similar mobility will not be present in samples prepared using control extracts prepared from untransformed cells.

In the alternative, REMAP function is assessed by expressing the sequences encoding REMAP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen Corporation, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2 μg of an additional

plasmid containing sequences encoding a marker protein are co-transfected.

Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; CLONTECH), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties.

FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of REMAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding REMAP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Inc., Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding REMAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIX. Identification of REMAP Ligands

REMAP is expressed in a eukaryotic cell line such as CHO (Chinese Hamster Ovary) or HEK (Human Embryonic Kidney) 293 which have a good history of GPCR expression and which contain a wide range of G-proteins allowing for functional coupling of the expressed REMAP to downstream effectors. The transformed cells are assayed for activation of the expressed receptors in the presence of candidate ligands. Activity is measured by changes in intracellular second messengers, such as cyclic AMP or Ca^{2+}. These may be measured directly using standard methods well known in the art, or by the use of reporter gene assays in which a luminescent protein (e.g. firefly luciferase or green fluorescent protein) is under the transcriptional control of a promoter responsive to the stimulation of protein kinase C by the activated receptor (Milligan, G. et al. (1996) Trends Pharmacol. Sci. 17:235-237). Assay technologies are available for both of these second messenger systems to allow high

throughput readout in multi-well plate format, such as the adenylyl cyclase activation FlashPlate Assay (NEN Life Sciences Products), or fluorescent Ca^{2+} indicators such as Fluo-4 AM (Molecular Probes) in combination with the FLIPR fluorimetric plate reading system (Molecular Devices). In cases where the physiologically relevant second messenger pathway is not known, REMAP may be coexpressed with the G-proteins $G_{\alpha 15/16}$ which have been demonstrated to couple to a wide range of G-proteins (Offermanns, S. and M.I. Simon (1995) J. Biol. Chem. 270:15175-15180), in order to funnel the signal transduction of the REMAP through a pathway involving phospholipase C and Ca^{2+} mobilization. Alternatively, REMAP may be expressed in engineered yeast systems which lack endogenous GPCRs, thus providing the advantage of a null background for REMAP activation screening. These yeast systems substitute a human GPCR and G_{α} protein for the corresponding components of the endogenous yeast pheromone receptor pathway. Downstream signaling pathways are also modified so that the normal yeast response to the signal is converted to positive growth on selective media or to reporter gene expression (Broach, J.R. and J. Thorner (1996) Nature 384 (supp.):14-16). The receptors are screened against putative ligands including known GPCR ligands and other naturally occurring bioactive molecules. Biological extracts from tissues, biological fluids and cell supernatants are also screened.

Various modifications and variations of the described compositions, methods, and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. It will be appreciated that the invention provides novel and useful proteins, and their encoding polynucleotides, which can be used in the drug discovery process, as well as methods for using these compositions for the detection, diagnosis, and treatment of diseases and conditions. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of embodiments within the scope of the invention. It is intended that the scope of the invention be defined by the following claims and their equivalents.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
2847449	1	2847449CD1	50	2847449CB1	2847449CA2, 6700829CA2
7523642	2	7523642CD1	51	7523642CB1	7724021CA2, 95104316CA2, 95104665CA2
7521994	3	7521994CD1	52	7521994CB1	95147645CA2
7522289	4	7522289CD1	53	7522289CB1	95156008CA2, 95156064CA2, 95156096CA2, 95156256CA2
7522336	5	7522336CD1	54	7522336CB1	90165149CA2, 95154989CA2
7522339	6	7522339CD1	55	7522339CB1	95156150CA2
7522361	7	7522361CD1	56	7522361CB1	95156142CA2
7522368	8	7522368CD1	57	7522368CB1	95154825CA2
7522373	9	7522373CD1	58	7522373CB1	95154909CA2, 95155157CA2, 95156755CA2
7522381	10	7522381CD1	59	7522381CB1	95155125CA2
7523596	11	7523596CD1	60	7523596CB1	95104072CA2
7523643	12	7523643CD1	61	7523643CB1	95088313CA2
7523769	13	7523769CD1	62	7523769CB1	95104279CA2
7523785	14	7523785CD1	63	7523785CB1	95104789CA2
7523836	15	7523836CD1	64	7523836CB1	95126325CA2
7523879	16	7523879CD1	65	7523879CB1	95092360CA2
7523880	17	7523880CD1	66	7523880CB1	
7523812	18	7523812CD1	67	7523812CB1	
7524026	19	7524026CD1	68	7524026CB1	
7524357	20	7524357CD1	69	7524357CB1	95035939CA2
7524808	21	7524808CD1	70	7524808CB1	
7522161	22	7522161CD1	71	7522161CB1	
7523999	23	7523999CD1	72	7523999CB1	90190792CA2
7524024	24	7524024CD1	73	7524024CB1	90170907CA2, 90171015CA2
7522455	25	7522455CD1	74	7522455CB1	90041650CA2, 90116608CA2, 90116672CA2, 90116680CA2, 90116716CA2, 90116763CA2, 90116808CA2, 90116856CA2, 90116880CA2, 90116888CA2
7524494	26	7524494CD1	75	7524494CB1	
7524965	27	7524965CD1	76	7524965CB1	7143112CA2, 95146918CA2

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
7525018	28	7525018CD1	77	7525018CB1	
7516620	29	7516620CD1	78	7516620CB1	
7525149	30	7525149CD1	79	7525149CB1	95147749CA2
7513047	31	7513047CD1	80	7513047CB1	
7513056	32	7513056CD1	81	7513056CB1	3450627CA2
7513245	33	7513245CD1	82	7513245CB1	
7513711	34	7513711CD1	83	7513711CB1	90072372CA2
7513965	35	7513965CD1	84	7513965CB1	95004626CA2
7513969	36	7513969CD1	85	7513969CB1	95016449CA2
7512119	37	7512119CD1	86	7512119CB1	90009957CA2
7515577	38	7515577CD1	87	7515577CB1	95040059CA2
7514748	39	7514748CD1	88	7514748CB1	90212745CA2
7513838	40	7513838CD1	89	7513838CB1	90043551CA2
7515163	41	7515163CD1	90	7515163CB1	95036620CA2, 95036660CA2, 95036676CA2, 95036684CA2, 95036844CA2, 95036928CA2, 95036944CA2, 95036952CA2
7516929	42	7516929CD1	91	7516929CB1	
7515570	43	7515570CD1	92	7515570CB1	90211864CA2, 95055714CA2, 95055746CA2, 95055894CA2
7515680	44	7515680CD1	93	7515680CB1	95041633CA2
7516698	45	7516698CD1	94	7516698CB1	90151102CA2, 90151134CA2
7517501	46	7517501CD1	95	7517501CB1	
7518576	47	7518576CD1	96	7518576CB1	95075765CA2
7518626	48	7518626CD1	97	7518626CB1	95073868CA2
7515714	49	7515714CD1	98	7515714CB1	95039247CA2

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
3	7521994CD1	g2913995	5.8E-155	[Homo sapiens] CD33L1
				Takei, Y. et al., Molecular cloning of a novel gene similar to myeloid antigen CD33 and its specific expression in placenta, Cytogenet. Cell Genet. 78, 295-300 (1997)
		334526|SIGLEC6	3.2E-147	[Homo sapiens][Cytoplasmic; Unspecified membrane; Plasma membrane] Sialic acid binding Ig-like lectin 6 (CD33 antigen-like), member of the sialoadhesin subgroup of the Ig superfamily, binds ligands containing Neu5Acalpha2-6GalNAcalpha (sialyl-Tn), may regulate serum leptin levels and play a role in cell-cell adhesion
				Brinkman-Van der Linden, E. C. et al., Loss of N-glycolylneuraminic acid in human evolution. Implications for sialic acid recognition by siglecs., J Biol Chem 275, 8633-40. (2000).
		752561|SIGLEC L1	7.3E-73	[Homo sapiens] SIGLEC-like 1, a member of the sialic acid binding immunoglobulin-like lectins, lacks essential arginine that is necessary for sialic acid recognition, may act as a negative regulator of signaling in macrophages by functioning as an inhibitory receptor
				Angata, T. et al., A second uniquely human mutation affecting sialic acid biology., J Biol Chem 276, 40282-7. (2001).
4	7522289CD1	g3641527	4.5E-34	[Homo sapiens] low-density lipoprotein receptor-related protein 5
				Hey, P. J. et al., Cloning of a novel member of the low-density lipoprotein receptor family, Gene 216, 103-111 (1998)
		336282|LRP5	3.3E-35	[Homo sapiens][Receptor (protein translocation)][Plasma membrane] Low density lipoprotein receptor related protein 5, involved in Wnt signaling and skeletal development; gene mutations cause osteoporosis-pseudoglioma syndrome (homozygous), reduced bone mass (heterozygous) and may be associated with type 1 diabetes
				Little, R. D. et al., A Mutation in the LDL Receptor-Related Protein 5 Gene Results in the Autosomal Dominant High-Bone-Mass Trait., Am J Hum Genet 70, 11-9. (2002).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		581963|Lrp5	8.7E-35	[Mus musculus][Receptor (protein translocation)][Unspecified membrane; Plasma membrane] Low density lipoprotein receptor related protein 5, may play a role in skeletal development; gene mutations in human LRP5 cause osteoporosis-pseudoglioma syndrome (homozygous), reduced bone mass (heterozygous) and may be associated with type 1 diabetes
				Kato, M. et al., Cbfa1-independent decrease in osteoblast proliferation, osteopenia, and persistent embryonic eye vascularization in mice deficient in Lrp5, a Wnt coreceptor., J Cell Biol 157, 303-14. (2002).
5	7522336CD1	g7188567	8.2E-54	[Homo sapiens] lectin-like receptor F1
				Roda-Navarro, P. et al., Human KLRF1, a novel member of the killer cell lectin-like receptor gene family: molecular characterization, genomic structure, physical mapping to the NK gene complex and expression analysis, Eur. J. Immunol. 30, 568-576 (2000)
		476579|KLRF1	5.9E-55	[Homo sapiens][Receptor (signalling)][Plasma membrane] Killer cell lectin-like receptor subfamily F member 1, contains immunoreceptor tyrosine-based inhibitory motifs, cooperates with NKp46 (LY94) to activate NK cell, induces NK-mediated cytotoxicity and lysis of PHA-induced lymphoblasts and Ca2+ mobilization
				Vitale, M. et al., Identification of NKp80, a novel triggering molecule expressed by human NK cells., Eur J Immunol 31, 233-42. (2001).
6	7522339CD1	g3925599	7.2E-46	[Homo sapiens] DORA protein
				Bates, E. E. et al., CD40L activation of dendritic cells down-regulates DORA, a novel member of the immunoglobulin superfamily, Mol. Immunol. 35, 513-524 (1998)
		342442|GSF6	5.3E-47	[Homo sapiens][Receptor (signalling)][Plasma membrane] Immunoglobulin superfamily member 6, member of CD8 family of receptors, downregulated by CD40L activation of dendritic cells (DC), may function in myeloid lineage commitment and antigen uptake or DC homing or recirculation

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Bates, E. E. et al., The mouse and human IGSF6 (DORA) genes map to the inflammatory bowel disease 1 locus and are embedded in an intron of a gene of unknown function., Immunogenetics 52, 112-20. (2000).
		701147|Igsf6	1.3E-25	[Mus musculus] Immunoglobulin superfamily member 6, expressed only in cells of the immune system, especially macrophages; gene is located within an intron of another, broadly expressed gene
7	7522361CD1	g3925599	3.6E-93	[Homo sapiens] DORA protein
				Bates, E. E. et al. (1998) (supra)
		342442|IGSF6	2.5E-94	[Homo sapiens][Receptor (signalling)][Plasma membrane] Immunoglobulin superfamily member 6, member of CD8 family of receptors, downregulated by CD40L activation of dendritic cells (DC), may function in myeloid lineage commitment and antigen uptake or DC homing or recirculation
				Bates, E. E. et al.(2000) (supra)
		784992|Igsf6	2.0E-55	[Rattus norvegicus][Receptor (signalling)] Protein containing an immunoglobulin (Ig) domain, which may be involved in protein-protein and protein-ligand interactions
8	7522368CD1	g7188567	6.3E-96	[Homo sapiens] lectin-like receptor F1
				Roda-Navarro, P. et al. (supra)
		476579|KLRF1	4.6E-97	[Homo sapiens][Receptor (signalling)][Plasma membrane] Killer cell lectin-like receptor subfamily F member 1, contains immunoreceptor tyrosine-based inhibitory motifs, cooperates with NKp46 (LY94) to activate NK cell, induces NK-mediated cytotoxicity and lysis of PHA-induced lymphoblasts and Ca2+ mobilization
				Vitale, M. et al. (supra)
		336168|KLRB1	6.8E-22	[Homo sapiens][Receptor (signalling); Small molecule-binding protein][Plasma membrane] Killer cell lectin-like receptor subfamily B member 1, involved in inhibiting T-cell cytolysis, transendothelium migration, and cell proliferation, elevated expression is observed in Guillain Barre syndrome and multiple sclerosis

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Borsellino, G. et al., Phenotypic and functional properties of gamma delta T cells from patients with Guillain Barre syndrome., J Neuroimmunol 102, 199-207. (2000).
9	7522373CD1	g7188567	1.1E-26	[Homo sapiens] lectin-like receptor F1
				Roda-Navarro, P. et al. (supra)
		476579|KLRF1	8.0E-28	[Homo sapiens][Receptor (signalling)][Plasma membrane] Killer cell lectin-like receptor subfamily F member 1, contains immunoreceptor tyrosine-based inhibitory motifs, cooperates with NKp46 (LY94) to activate NK cell, induces NK-mediated cytotoxicity and lysis of PHA-induced lymphoblasts and Ca2+ mobilization
				Vitale, M. et al. (supra)
10	7522381CD1	g7188567	1.8E-26	[Homo sapiens] lectin-like receptor F1
				Roda-Navarro, P. et al. (supra)
		476579|KLRF1	1.3E-27	[Homo sapiens][Receptor (signalling)][Plasma membrane] Killer cell lectin-like receptor subfamily F member 1, contains immunoreceptor tyrosine-based inhibitory motifs, cooperates with NKp46 (LY94) to activate NK cell, induces NK-mediated cytotoxicity and lysis of PHA-induced lymphoblasts and Ca2+ mobilization
				Vitale, M. et al. (supra)
11	7523596CD1	g2062706	4.6E-163	[Homo sapiens] butyrophilin
				Ruddy, D. A. et al., A 1.1-Mb transcript map of the hereditary hemochromatosis locus, Genome Res. 7, 441-456 (1997)
		789729|BTN3A1	3.3E-164	[Homo sapiens] Butyrophilin subfamily 3A1, a member of the butyrophilin family, which are components of the milk fat globule membrane, contains immunoglobulin variable and constant domains
				Rhodes, D. A. et al., The cluster of btn genes in the extended major histocompatibility complex., Genomics 71, 351-62. (2001).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		428258|BTN3A2	5.5E-139	[Homo sapiens] Butyrophilin subfamily 3A2, a member of the butyrophilin family, which are components of the milk fat globule membrane, contains immunoglobulin variable and constant domains
12	7523643CD1	g887941	5.1E-196	[Mus musculus] pmel17 protein
				Kwon, B. S. et al., Mouse silver mutation is caused by a single base insertion in the putative cytoplasmic domain of Pmel 17, Nucleic Acids Res. 23, 154-158 (1995)
		626518|Si	3.7E-197	[Mus musculus] Melanosomal protein silver, plays a role in melanogenesis and determination of coat color phenotype, a tumor antigen and immunotherapeutic target for melanoma
				Sturm, R. A. et al. Human pigmentation genes: identification, structure and consequences of polymorphic variation. Gene 277, 49-62. (2001).
		432514|SILV	1.6E-156	[Homo sapiens][Golgi; Endosome/Endosomal vesicles; Cytoplasmic; Extracellular (excluding cell wall); Plasma membrane] Melanosomal matrix protein silver-like, predicted to play a role in melanogenesis, is a likely candidate gene for some forms of oculocutaneous albinism, a frequent tumor antigen and immunotherapeutic target for metastatic melanoma
				Wagner, S. N. et al., Analysis of Pmel17/gp100 expression in primary human tissue specimens: implications for melanoma immuno- and gene-therapy., Cancer Immunol Immunother 44, 239-47 (1997).
13	7523769CD1	g2062706	3.0E-72	[Homo sapiens] butyrophilin
				Ruddy, D. A. et al. (supra)
		789729|BTN3A1	2.2E-73	[Homo sapiens] Butyrophilin subfamily 3A1, a member of the butyrophilin family, which are components of the milk fat globule membrane, contains immunoglobulin variable and constant domains
				Rhodes, D. A. et al. (supra)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		339696|BTN1A1	4.2E-34	[Homo sapiens][Receptor (signalling)][Plasma membrane] Butyrophilin subfamily 1 member A1, a putative integral membrane protein of the immunoglobulin superfamily, secreted in association with the milk-fat-globule membrane from mammary epithelial cells
				Taylor, M. R. et al., Cloning and sequence analysis of human butyrophilin reveals a potential receptor function., Biochim Biophys Acta 1306, 1-4 (1996)
14	7523785CD1	g2062706	6.2E-152	[Homo sapiens] butyrophilin
				Ruddy, D.A. et al. A 1.1-Mb transcript map of the hereditary hemochromatosis locus. Genome Res. 7:441-456 (1997).
		789729|BTN3A1	4.4E-153	[Homo sapiens] Butyrophilin subfamily 3A1, a member of the butyrophilin family, which are components of the milk fat globule membrane, contains immunoglobulin variable and constant domains
				Rhodes, D.A. et al. The cluster of btn genes in the extended major histocompatibility complex. Genomics 71:351-362 (2001).
		586905|Btn1a1	4.9E-81	[Mus musculus] [Receptor (signaling)] [Plasma membrane; Extracellular (excluding cell wall)] Butyrophilin subfamily 1 member A1, a putative integral membrane protein of the immunoglobulin superfamily, secreted in association with the milk-fat-globule membrane from mammary epithelial cells
				Ogg, S.L. et al. Structural organization and mammary-specific expression of the butyrophilin gene. Mamm. Genome 7:900-905 (1996).
				Banghart, L.R. et al. Butyrophilin is expressed in mammary epithelial cells from a single-sized messenger RNA as a type I membrane glycoprotein. J. Biol. Chem. 273:4171-4179 (1998).
15	7523836CD1	g402207	2.5E-288	[Homo sapiens] endoglin
				Bellon, T. et al. Identification and expression of two forms of the human transforming growth factor-beta-binding protein endoglin with distinct cytoplasmic regions. Eur. J. Immunol. 23:2340-2345 (1993).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		339318|ENG	1.8E-289	[Homo sapiens] [Receptor (signaling); Small molecule-binding protein] [Plasma membrane] Endoglin, an integral membrane glycoprotein that binds transforming growth factor beta 1 (TGFB1) and plays a critical role in angiogenesis and heart development; mutations in the corresponding gene are associated with hereditary hemorrhagic telangiectasia
				McAllister, K.A. et al. Endoglin, a TGF-beta binding protein of endothelial cells, is the gene for hereditary haemorrhagic telangiectasia type 1. Nat. Genet. 8:345-351 (1994).
				Abdalla, S.A. et al. Analysis of ALK-1 and endoglin in newborns from families with hereditary hemorrhagic telangiectasia type 2. Hum. Mol. Genet. 9:1227-1237 (2000).
				Barbara, N.P. et al. Endoglin is an accessory protein that interacts with the signaling receptor complex of multiple members of the transforming growth factor-beta superfamily. J. Biol. Chem. 274:584-594 (1999).
		582747|Eng	1.0E-189	[Mus musculus] [Unspecified membrane] Endoglin, an integral membrane glycoprotein that binds transforming growth factor-beta 1 (Tgfb1) and plays a critical role in angiogenesis and heart development; mutations in the human ENG gene are associated with hereditary hemorrhagic telangiectasia
				Li, D.Y. et al. Defective angiogenesis in mice lacking endoglin. Science 284:1534-1537 (1999).
				Arthur, H.M. et al. Endoglin, an ancillary TGFbeta receptor, is required for extraembryonic angiogenesis and plays a key role in heart development. Dev. Biol. 217:42-53 (2000).
				Bourdeau, A. et al. A murine model of hereditary hemorrhagic telangiectasia. J. Clin. Invest. 104:1343-1351 (1999).
16	7523879CD1	g3153241	0.0	[Homo sapiens] class I cytokine receptor
				Sprecher, C.A. et al. Cloning and characterization of a novel class I cytokine receptor. Biochem. Biophys. Res. Commun. 246:82-90 (1998).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		341448|WSX1	0.0	[Homo sapiens] [Receptor (signaling)] [Plasma membrane] Cytokine receptor family class 1 member 1, a putative transmembrane receptor that may be involved in regulation of the immune response; mouse Wsx1 is involved in the response to Leishmania major and mycobacterial infection
				Yoshida, H. et al. WSX-1 is required for the initiation of Th1 responses and resistance to L. major infection. Immunity 15:569-578 (2001).
		587585|Tccr	7.9E-131	[Mus musculus] [Receptor (signaling)] [Plasma membrane] Cytokine receptor family class 1 member 1, a protein involved in Th1 cell differentiation and IFN-gamma (Ifng) production, may also play a role in inhibition of T cell proliferation; involved in the response to Leishmania major and mycobacterial infection
				Chen, Q. et al. Development of Th1-type immune responses requires the type I cytokine receptor TCCR. Nature 407:916-920 (2000).
17	7523880CD1	g666043	1.8E-285	[Homo sapiens] NMB
				Weterman, M.A. et al. nmb, a novel gene, is expressed in low-metastatic human melanoma cell lines and xenografts. Int. J. Cancer 60:73-81 (1995).
		342930|GPNMB	1.3E-286	[Homo sapiens] Glycoprotein (transmembrane) nmb, has similarity to a melanocyte-specific protein pMEL17, has greater expression in low metastatic melanoma cell lines than in highly metastatic lines, may play a role in retardation of growth and attenuation of metastatis
				Weterman, M.A. et al., *supra*.
		753925|Gpnmb	2.2E-198	[Mus musculus] Glycoprotein (transmembrane) nmb, a dendritic cell (DC)-associated adhesion molecule that binds heparan sulfate proteoglycans on endothelial cells and may promote transendothelial migration of DC, also has an RGD sequence that may promote integrin binding
				Shikano, S. et al. Molecular cloning of a dendritic cell-associated transmembrane protein, DC-HIL, that promotes RGD-dependent adhesion of endothelial cells through recognition of heparan sulfate proteoglycans. J. Biol. Chem. 276:8125-8134 (2001).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
18	7523812CD1	g2281008	2.6E-73	[Homo sapiens] Eph-family protein
				Matsuoka, H. et al. Expression of a kinase-defective Eph-like receptor in the normal human brain. Biochem. Biophys. Res. Commun. 235:487-492 (1997).
		340488|EPHB6	1.9E-74	[Homo sapiens] [Protein kinase; Transferase; Inhibitor or repressor; Receptor (signaling)] [Plasma membrane] Eph-related receptor tyrosine kinase B6 (EPHB6), putative catalytically-inactive receptor tyrosine kinase, binds ephrin-B2 (EFNB2), may regulate T cell development, cytokine production, and apoptosis, upregulation correlates with reduced tumorigenicity
				Tang, X.X. et al. Implications of EPHB6, EFNB2, and EFNB3 expressions in human neuroblastoma. Proc. Natl. Acad. Sci. USA 97:10936-10941 (2000).
				Shimoyama, M. et al. T-cell-specific expression of kinase-defective Eph-family receptor protein, EphB6 in normal as well as transformed hematopoietic cells. Growth Factors 18:63-78 (2000).
				Luo, H. et al. Cross-linking of EphB6 resulting in signal transduction and apoptosis in Jurkat cells. J. Immunol. 167:1362-1370 (2001).
		581027|Ephb6	1.9E-70	[Mus musculus] [Protein kinase; Transferase; Receptor (signaling)] [Plasma membrane] Eph-related receptor tyrosine kinase B6 (EphB6), putative catalytically-inactive receptor tyrosine kinase that may bind ephrin-B2 (Efnb2), may be involved in T cell development; elevated human EPHB6 gene expression correlates with reduced tumorigenicity
				Munthe, E. et al. Ephrin-B2 is a candidate ligand for the Eph receptor, EphB6. FEBS Lett. 466:169-174 (2000).
19	7524026CD1	g31960	1.0E-133	[Homo sapiens] G-Tsf precursor
				de Martin, R. et al. Complementary DNA for human glioblastoma-derived T cell suppressor factor, a novel member of the transforming growth factor-beta gene family. EMBO J. 6:3673-3677 (1987).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		338480|TGFB2	7.5E-135	[Homo sapiens] [Ligand; Inhibitor or repressor] [Extracellular (excluding cell wall)] Transforming growth factor beta 2, a member of the transforming growth factor beta superfamily, regulates the immune response; elevated in plasma of colon carcinoma patient and may play a role in cerebral malaria pathogenesis
				Webb, N.R. et al. Structural and sequence analysis of TGF-beta 2 cDNA clones predicts two different precursor proteins produced by alternative mRNA splicing. DNA 7:493-497 (1988).
		581643|Tgfb2	3.5E-128	[Mus musculus] [Ligand; Inhibitor or repressor] [Extracellular (excluding cell wall)] Transforming growth factor beta 2, a member of the transforming growth factor beta superfamily, mediates growth suppression and may regulate the immune response
				Shipley, J.M. et al. Developmental expression of latent transforming growth factor beta binding protein 2 and its requirement early in mouse development. Mol. Cell. Biol. 20:4879-4887 (2000).
				Robinson, S.D. et al. Regulated expression and growth inhibitory effects of transforming growth factor-beta isoforms in mouse mammary gland development. Dev. Suppl. 113:867-878 (1991).
20	7524357CD1	g12859694	2.2E-55	[Mus musculus] homolog to BRUSH BORDER 61.9 KDA PROTEIN PRECURSOR
21	7524808CD1	g2062702	1.1E-216	[Homo sapiens] butyrophilin
				Ruddy, D. A. et al., A 1.1-Mb transcript map of the hereditary hemochromatosis locus, Genome Res. 7, 441-456 (1997)
		568298|BTN2A2	7.5E-218	[Homo sapiens] Butyrophilin 2, member of the butyrophilin family, which is a subset of the immunoglobulin superfamily, with immunoglobulin, transmembrane, and B30.2 domains
				Rhodes, D. A. et al., The cluster of btn genes in the extended major histocompatibility complex., Genomics 71, 351-62. (2001).
		762563|BTN2A1	1.1E-64	[Homo sapiens][Plasma membrane] Butyrophilin subfamily 2A1, a member of the butyrophilin family, contains two immunoglobulin domains

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Tazi-Ahnini, R. et al., Cloning, localization, and structure of new members of the butyrophilin gene family in the juxta-telomeric region of the major histocompatibility complex., Immunogenetics 47, 55-63 (1997).
22	7522161CD1	g726513	1.1E-126	[Homo sapiens] nuclear orphan receptor LXR-alpha
				Willy, P. J. et al., LXR, a nuclear receptor that defines a distinct retinoid response pathway, Genes Dev. 9, 1033-1045 (1995)
		342610|NR1H3	7.4E-128	[Homo sapiens][Activator; Transcription factor; DNA-binding protein; Receptor (signalling)][Nuclear] Nuclear receptor subfamily 1 group H member 3, a ligand-dependent nuclear receptor transcription factor, acts in RXR receptor and PPARA signaling, cholesterol/lipid metabolism, and TNF synthesis; mutations in mouse Nr1h3 result in cholesterol accumulation
				Costet, P. et al., Sterol-dependent transactivation of the ABC1 promoter by the liver X receptor/retinoid X receptor., J Biol Chem 275, 28240-5 (2000).
		587195|Nr1h3	1.1E-108	[Mus musculus][Activator; DNA-binding protein; Transcription factor; Receptor (signalling)][Nuclear] Nuclear receptor subfamily 1 group H member 3, a ligand-dependent nuclear receptor transcription factor, acts in RXR receptor and Pparg signaling as well as cholesterol, fatty acid, and bile metabolism; mutations result in cholesterol accumulation
				Wan, Y. J. et al., Hepatocyte-specific mutation establishes retinoid X receptor alpha as a heterodimeric integrator of multiple physiological processes in the liver., Mol Cell Biol 20, 4436-44 (2000).
23	7523999CD1	g6164831	0.0	[Rattus norvegicus] Robo2
				Brose, K. et al., Slit proteins bind Robo receptors and have an evolutionarily conserved role in repulsive axon guidance, Cell 96, 795-806 (1999)
24	7524024CD1	g25137569	0.0	NPEH2 [Homo sapiens]
				Donoviel, D. B. et al., Proteinuria and perinatal lethality in mice lacking NEPH1, a novel protein with homology to NEPHRIN, Mol. Cell. Biol. 21, 4829-4836 (2001)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		735866|KIAA1867	0.0	[Homo sapiens] Protein containing five immunoglobulin (Ig) domains, which may be involved in protein-protein and protein-ligand interactions, has a region of low similarity to a region of nephrosis 1 (mouse Nphs1), which may have a role in cell-cell interactions
25	7522455CD1	g339887	2.6E-139	[Homo sapiens] steroid receptor TR2-11
				Chang, C. et al., Molecular cloning of new human TR2 receptors: a class of steroid receptor with multiple ligand-binding domains, Biochem. Biophys. Res. Commun. 165, 735-741 (1989)
		338676|NR2C1	1.8E-140	[Homo sapiens][Inhibitor or repressor; Transcription factor; DNA-binding protein; Receptor (signalling)][Nuclear] Nuclear receptor subfamily 2 group C member 1, a member of nuclear hormone receptor transcription factor family, functions both as a transcriptional activator and repressor; multiple alternative forms with different ligand-binding domains are detected
				Collins, L. L. et al., Feedback regulation between orphan nuclear receptor TR2 and human papilloma virus type 16., J Biol Chem 276, 27316-21. (2001).
		586455|Nr2c1	5.0E-122	[Mus musculus][Inhibitor or repressor; DNA-binding protein; Transcription factor; Receptor (signalling)][Nuclear] Nuclear receptor subfamily 2 group C member 1, functions both as a transcriptional activator and repressor, may play role in spermatogenesis
				Lee, C. H. et al., Characterization of the mouse nuclear orphan receptor TR2-11 gene promoter and its potential role in retinoic acid-induced P19 apoptosis., Biochem Pharmacol 60, 127-36 (2000).
				Franco, P. J. et al., The orphan nuclear receptor TR2 interacts directly with both class I and class II histone deacetylases., Mol Endocrinol 15, 1318-28. (2001).
26	7524494CD1	g36110	0.0	[Homo sapiens] tyrosine kinase

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		581203|Mst1r	0.0	[Mus musculus][Protein kinase; Transferase; Receptor (signalling)][Plasma membrane] Macrophage-stimulating protein receptor, a receptor tyrosine kinase that regulates inflammatory responses, cell migration and IFN-gamma (Ifng) -induced nitric oxide production by macrophages, inhibits apoptosis; may play a role in Friend SFFV infection
				Kurihara, N. et al., Macrophage-stimulating protein activates STK receptor tyrosine kinase on osteoclasts and facilitates bone resorption by osteoclast-like cells., Blood 87, 3704-10 (1996).
		429302|Met	8.6E-205	[Mus musculus][Protein kinase; Transferase; Receptor (signalling)][Plasma membrane] Met protooncogene, member of the receptor tyrosine kinase family of oncogenes, functions in cell proliferation, differentiation, and migration and in organ development; acts as a receptor for a Listeria monocytogenes surface protein
				Chan, A. M. et al., Characterization of the mouse met proto-oncogene., Oncogene 2, 593-9 (1988).
27	7524965CD1	g32064	3.9E-23	[Homo sapiens] hepsin
				Leytus, S. P. et al., A novel trypsin-like serine protease (hepsin) with a putative transmembrane domain expressed by human liver and hepatoma cells, Biochemistry 27, 1067-1074 (1988)
		335864|HPN	2.7E-24	[Homo sapiens][Hydrolase; Protease (other than proteasomal)][Plasma membrane] Hepsin, a transmembrane serine protease implicated in cell growth control and initiation of blood coagulation; overexpressed in prostate and ovarian tumors
				Dhanasekaran, S. M. et al., Delineation of prognostic biomarkers in prostate cancer., Nature 412, 822-6. (2001).
		591137|Hpn	6.1E-19	[Rattus norvegicus][Hydrolase; Protease (other than proteasomal)][Plasma membrane] Hepsin, a transmembrane serine protease; human HPN is overexpressed in prostate and ovarian tumors

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Zhukov, A. et al., Purification and characterization of hepsin from rat liver microsomes., Biochim Biophys Acta 1337, 85-95. (1997).
28	7525018CD1	g507151	0.0	[Homo sapiens] IL12 receptor component
		342530|IL12RB1	0.0	[Homo sapiens][Receptor (signalling)][Plasma membrane] Interleukin 12 receptor-beta 1, subunit of the IL12 receptor, mediates IL12-induced Th1 cell response to pathogens, T-cell proliferation, induction of IFNG; mutation in corresponding gene results in susceptibility to mycobacteria and salmonella infections
				Oppmann, B. et al., Novel p19 protein engages IL-12p40 to form a cytokine, IL-23, with biological activities similar as well as distinct from IL-12, Immunity 13, 715-25 (2000).
		583351|Il12rb1	3.3E-164	[Mus musculus][Receptor (signalling)][Plasma membrane] Interleukin 12 receptor-beta 1, subunit of the IL12 receptor, mediates IL12-induced Th1 cell response to intracellular pathogens, T-cell proliferation, induction of IFNG and natural killer cell cytotoxicity, enhances immune response against allografts
				Salkowski, C. A. et al., Impaired IFN-gamma production in IFN regulatory factor-1 knockout mice during endotoxemia is secondary to a loss of both IL-12 and IL-12 receptor expression, J Immunol 165, 3970-7 (2000).
29	7516620CD1	g1200235	0.0	[Homo sapiens] SEX protein
		582525|Plxn1	0.0	[Mus musculus][Protein phosphatase; Hydrolase; Receptor (signalling)][Plasma membrane] Plexin 1, acts as a coreceptor with neuropilin (Nrp) for semaphorins, which are signaling molecules controlling cell repulsion
		741907|PLXNA1	0.0	[Homo sapiens] Plexin A1, acts as a coreceptor with neuropilin (NRP1) for semaphorins, which are signaling molecules controlling cell repulsion
				Tamagnone, L. et al., Plexins are a large family of receptors for transmembrane, secreted, and GPI-anchored semaphorins in vertebrates., Cell 99, 71-80 (1999).
30	7525149CD1	g15281418	0.0	[Homo sapiens] keratinocyte growth factor receptor 2 isoform K-sam-IIC2
				Ingersoll, R. G. et al., Fibroblast growth factor receptor 2 (FGFR2): genomic sequence and variations, Cytogenet. Cell Genet. 94, 121-126 (2001)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		690562|FGFR2	0.0	[Homo sapiens][Protein kinase; Transferase; Receptor (signalling)][Plasma membrane; Extracellular (excluding cell wall)] Fibroblast growth factor receptor 2, receptor tyrosine kinase that binds fibroblast growth factors, plays roles in proliferation and development; mutations are associated with autosomal dominant craniosynostotic syndromes
				Anderson, J. et al., Apert syndrome mutations in fibroblast growth factor receptor 2 exhibit increased affinity for FGF ligand., Hum Mol Genet 7, 1475-83 (1998).
		584789|Fgfr2	0.0	[Mus musculus][Protein kinase; Transferase; Receptor (signalling); Small molecule-binding protein][Plasma membrane] Fibroblast growth factor receptor 2, receptor tyrosine kinase that binds fibroblast growth factors and plays roles in proliferation, development and apoptosis; mutations in human FGFR2 are associated with craniosynostotic syndromes
				Twigg, S. R. et al., Conserved use of a non-canonical 5' splice site (/GA) in alternative splicing by fibroblast growth factor receptors 1, 2 and 3., Hum Mol Genet 7, 685-91 (1998).
31	7513047CD1	g3171912	1.9E-23	[Homo sapiens] RAMP2
				McLatchie, L.M. et al. RAMPs regulate the transport and ligand specificity of the calcitonin-receptor-like receptor. Nature 393:333-339 (1998).
		342728|RAMP2	1.5E-24	[Homo sapiens] [Regulatory subunit] [Lysosome/vacuole; Endosome/Endosomal vesicles; Cytoplasmic; Plasma membrane] Receptor activity-modifying protein 2, an adrenomedullin (ADM) receptor in complex with calcitonin receptor-like receptor (CALCRL); involved in endotoxemia response in mouse lung and obstructive nephropathy and ischemic heart failure in rats
				Kamitani, S. et al. The RAMP2/CRLR complex is a functional adrenomedullin receptor in human endothelial and vascular smooth muscle cells. FEBS Lett. 448:111-114 (1999).
				Husmann, K. et al. Mouse receptor-activity-modifying proteins 1, -2 and -3: amino acid sequence, expression and function. Mol. Cell. Endocrinol. 162:35-43 (2000).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
32	7513056CD1	g703118	2.8E-29	[Homo sapiens] thyroid receptor interactor
				Lee, J.W. et al. Two classes of proteins dependent on either the presence or absence of thyroid hormone for interaction with the thyroid hormone receptor. Mol. Endocrinol. 9:243-254 (1995).
		344632|NFKBIB	2.2E-30	[Homo sapiens] Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor beta (I kappa B-beta), an isoform of I kappa B that inhibits the transcription factor NF-kappa B
				Bourcier, T. et al. The nuclear factor kappa B signaling pathway participates in dysregulation of vascular smooth muscle cells in vitro and in human atherosclerosis. J. Biol. Chem. 272:15817-15824 (1997).
				Prigent, M. et al. I kappa B-alpha and I kappa B-alpha /NF-kappa B complexes are retained in the cytoplasm through interaction with a novel partner, rasGAP SH3-binding protein 2. J. Biol. Chem. 275:36441-36449 (2000).
		585393|Nfkbib	2.0E-29	[Mus musculus] [Small molecule-binding protein] I kappa B-beta, an isoform of I kappa B that plays a role in the regulation of NF-kappa B
				Thompson, J.E. et al. I kappa B-beta regulates the persistent response in a biphasic activation of NF-kappa B. Cell 80:573-582 (1995).
				Velasco, M. et al. Rapid up-regulation of I kappa B-beta and abrogation of NF-kappa B activity in peritoneal macrophages stimulated with lipopolysaccharide. J. Biol. Chem. 272:23025-23030 (1997).
33	7513245CD1	g2344802	7.8E-169	[Homo sapiens] ApoER2
				Kim, D.H. et al. Exon/intron organization, chromosome localization, alternative splicing, and transcription units of the human apolipoprotein E receptor 2 gene. J. Biol. Chem. 272:8498-8504 (1997).
		749048|LRP8	1.8E-145	[Homo sapiens] [Receptor (signaling)] [Plasma membrane] Apolipoprotein E receptor 2 (low density lipoprotein receptor-related protein 8), a low density lipoprotein receptor, may play roles in signal transduction and cell migration, may be involved in receptor-mediated endocytosis

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Trommsdorff, M. et al. Reeler/Disabled-like disruption of neuronal migration in knockout mice lacking the VLDL receptor and ApoE receptor 2. Cell 97:689-701 (1999).
				Sun, X.M. et al. Expression in vitro of alternatively spliced variants of the messenger RNA for human apolipoprotein E receptor-2 identified in human tissues by ribonuclease protection assays. Eur. J. Biochem. 262:230-239 (1999).
				Clatworthy, A.E. et al. Expression and alternate splicing of apolipoprotein E receptor 2 in brain. Neuroscience 90:903-911 (1999).
		753863|Lrp8	7.9E-145	[Mus musculus] [Apical plasma membrane] Apolipoprotein E receptor 2 (low density lipoprotein receptor-related protein 8), a likely low density lipoprotein receptor, may play roles in signal transduction and neuronal cell migration, may be involved in cytoskeleton organization and biogenesis
				Stockinger, W. et al. The reelin receptor ApoER2 recruits JNK-interacting proteins-1 and -2. J. Biol. Chem. 275:25625-25632 (2000).
				Hiesberger, T. et al. Direct binding of Reelin to VLDL receptor and ApoE receptor 2 induces tyrosine phosphorylation of disabled-1 and modulates tau phosphorylation. Neuron 24:481-489 (1999).
34	7513711CD1	g1695876	7.1E-40	[Homo sapiens] interleukin-13 receptor
				Aman, M.J. et al. cDNA cloning and characterization of the human interleukin-13 receptor alpha chain. J. Biol. Chem. 271:29265-29270 (1996).
		335988|IL13RA1	5.6E-41	[Homo sapiens] [Receptor (signaling)] [Plasma membrane] Interleukin-13 receptor chain alpha, binds interleukin-13 (IL-13) with a low affinity, associates with IL-4R alpha (IL4R) and may also be a component of IL-4 receptors; mutation is associated with high levels of IgE, which may promote atopy
				Miloux, B. et al. Cloning of the human IL-13R alpha1 chain and reconstitution with the IL4R alpha of a functional IL-4/IL-13 receptor complex. FEBS Lett. 401:163-166 (1997).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Murata, T. et al. Structure of IL-13 receptor: analysis of subunit composition in cancer and immune cells. Biochem. Biophys. Res. Commun. 238:90-94 (1997).
				Heinzmann, A. et al. Genetic variants of IL-13 signalling and human asthma and atopy. Hum. Mol. Genet. 9:549-559 (2000).
35	7513965CD1	g1200070	4.3E-65	[Homo sapiens] interleukin 2 receptor
				Ishida, N. et al. Molecular cloning and structure of the human interleukin 2 receptor gene. Nucleic Acids Res. 13:7579-7589 (1985).
		339430|IL2RA	3.4E-66	[Homo sapiens] [Receptor (signaling)] [Plasma membrane] Alpha subunit of the interleukin-2 (IL2) receptor, component of T cell mediated immune response; deficiency causes autoimmune disorders, immunodeficiency, and abnormal T cell proliferation
				Leonard, W.J. et al. Localization of the gene encoding the human interleukin-2 receptor on chromosome 10. Science 228:1547-1549 (1985).
				Sharfe, N. et al. Human immune disorder arising from mutation of the alpha chain of the interleukin-2 receptor. Proc. Natl. Acad. Sci. USA 94:3168-3171 (1997).
				Horiuchi, S. et al. Altered interleukin-2 receptor alpha-chain is expressed in human T-cell leukemia virus type-I-infected T-cell lines and human peripheral blood mononuclear cells of adult T-cell leukemia patients through an alternative splicing mechanism. Immunology 91:28-34 (1997).
		590829|Il2ra	3.8E-35	[Rattus norvegicus] [Receptor (signaling)] [Plasma membrane] Alpha subunit of the interleukin-2 receptor, probable component of immune and inflammatory responses; deficiency of human IL2RA causes autoimmune disorders and immunodeficiency
				Page, T.H. et al. Molecular cloning of cDNAs for the rat interleukin 2 receptor alpha and beta chain genes: differentially regulated gene activity in response to mitogenic stimulation. Eur. J. Immunol. 21:2133-2138 (1991).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Wang, G. et al. Immunohistochemical localization of interleukin-2 and its receptor subunits alpha, beta and gamma in the main olfactory bulb of the rat. Brain Res. 893:244-252 (2001).
36	7513969CD1	g339665	9.9E-244	[Homo sapiens] thyroid hormone receptor
				Nakai, A. et al. Characterization of a thyroid hormone receptor expressed in human kidney and other tissues. Proc. Natl. Acad. Sci. USA. 85:2781-2785 (1988).
		338510|THRA	8.9E-244	[Homo sapiens] [Activator; Transcription factor; DNA-binding protein; Receptor (signaling)] [Nuclear] Thyroid hormone receptor alpha, a member of a family of ligand-activated nuclear transcription factors, a high affinity receptor for thyroid hormone that binds DNA and regulates a number of physiological processes
				Nagaya, T. et al. Heterodimerization preferences of thyroid hormone receptor alpha isoforms. Biochem. Biophys. Res. Commun. 226:426-430 (1996).
				Sap, J. et al. The c-erb-A protein is a high-affinity receptor for thyroid hormone. Nature 324:635-640 (1986).
		705060|Thra1	6.7E-237	[Rattus norvegicus] [Activator; Transcription factor; DNA-binding protein; Receptor (signaling)] [Nuclear] Thyroid hormone receptor alpha, a member of a family of ligand-activated nuclear transcription factors, a high affinity receptor for thyroid hormone that binds DNA and regulates a number of physiological processes
				Thompson, C.C. et al. Identification of a novel thyroid hormone receptor expressed in the mammalian central nervous system. Science 237:1610-1614 (1987).
				Tuca, A. et al. Ontogeny of thyroid hormone receptors and c-erbA expression during brown adipose tissue development: evidence of fetal acquisition of the mature thyroid status. Endocrinology 132:1913-1920 (1993).
39	7514748CD1	g1881447	1.0E-15	[Homo sapiens] coxsackie and adenovirus receptor protein
				Bergelson, J.M. et al. Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. Science 275:1320-1323 (1997).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		334918|CXADR	8.2E-17	[Homo sapiens] [Small molecule-binding protein] [Plasma membrane] Coxsackievirus and adenovirus receptor, transmembrane receptor that mediates viral attachment to the cell surface and intracellular viral uptake, involved in regulation of cell proliferation, may play a role in the progression of bladder cancer
				Bergelson, J.M. et al. (1997), *supra.*
				Tomko, R.P. et al. HCAR and MCAR: the human and mouse cellular receptors for subgroup C adenoviruses and group B coxsackieviruses. Proc. Natl. Acad. Sci. USA 94:3352-3356 (1997).
				Bowles, K.R. et al. Genomic organization and chromosomal localization of the human Coxsackievirus B-adenovirus receptor gene. Hum. Genet. 105:354-359 (1999).
		724666|1eaj_A	8.3E-17	[Protein Data Bank] Coxsackie Virus and Adenovirus Receptor
		732931|Cxadr	2.1E-16	[Rattus norvegicus] Coxsackievirus and adenovirus receptor, transmembrane receptor that mediates viral attachment to the cell surface and intracellular viral uptake, may play a role in cardiomyocyte adhesion; expressed highly during experimental autoimmune myocarditis
				Ito, M. et al. Expression of coxsackievirus and adenovirus receptor in hearts of rats with experimental autoimmune myocarditis. Circ. Res. 86:275-280 (2000).
		725033|1kac_B	7.5E-16	[Protein Data Bank] Coxsackie Virus and Adenovirus Receptor
40	7513838CD1	g2281008	2.3E-266	[Homo sapiens] Eph-family protein
				Matsuoka, H. et al. Expression of a kinase-defective Eph-like receptor in the normal human brain. Biochem. Biophys. Res. Commun. 235:487-492 (1997).
		340488|EPHB6	1.8E-267	[Homo sapiens] [Protein kinase; Transferase; Inhibitor or repressor; Receptor (signaling)] [Plasma membrane] Eph-related receptor tyrosine kinase B6 (EPHB6), putative catalytically-inactive receptor tyrosine kinase, binds ephrin-B2 (EFNB2), may regulate T cell development, cytokine production, and apoptosis, upregulation correlates with reduced tumorigenicity

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Tang, X.X. et al. Implications of EPHB6, EFNB2, and EFNB3 expressions in human neuroblastoma. Proc. Natl. Acad. Sci. USA 97:10936-10941 (2000).
				Shimoyama, M. et al. T-cell-specific expression of kinase-defective Eph-family receptor protein, EphB6 in normal as well as transformed hematopoietic cells. Growth Factors 18:63-78 (2000).
				Munthe, E. et al. Ephrin-B2 is a candidate ligand for the Eph receptor, EphB6. FEBS Lett. 466:169-174 (2000).
				Luo, H. et al. Cross-linking of EphB6 resulting in signal transduction and apoptosis in Jurkat cells. J. Immunol. 167:1362-1370 (2001).
		581027|Ephb6	2.1E-244	[Mus musculus] [Protein kinase; Transferase; Receptor (signaling)] [Plasma membrane] Eph-related receptor tyrosine kinase B6 (EphB6), putative catalytically-inactive receptor tyrosine kinase that may bind ephrin-B2 (Efnb2), may be involved in T cell development; elevated human EPHB6 gene expression correlates with reduced tumorigenicity
				Gurniak, C.B. et al. A new member of the Eph family of receptors that lacks protein tyrosine kinase activity. Oncogene 13:777-786 (1996).
				Shimoyama, M. et al. (2000), *supra.*
				Munthe, E. et al. (2000), *supra.*
41	7515163CD1	g15808375	3.4E-99	[Homo sapiens] erythroid membrane-associated protein
				Xu, H. et al. Cloning and characterization of human erythroid membrane-associated protein, human ERMAP. Genomics 76:2-4 (2001).
		789851|ERMAP	2.6E-100	[Homo sapiens] Erythroblast membrane-associated protein, a member of the B30.2 domain protein family that may play a role in red blood cell differentiation
				Xu, H. et al. (2001), *supra.*
		438203|Ermap	3.9E-37	[Mus musculus][Receptor (signalling)][Plasma membrane] Erythroid membrane-associated protein, a putative transmembrane glycoprotein with two extracellular immunoglobulin folds and a B30.2 domain, may be an adhesion or receptor molecule

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Ye, T.Z. et al. Ermap, a gene coding for a novel erythroid specific adhesion/receptor membrane protein. Gene 242:337-345 (2000).
42	7516929CD1	g11935053	3.1E-167	[Homo sapiens] sarcolemmal associated protein 1
				Wielowieyski, P.A. et al. Alternative splicing, expression, and genomic structure of the 3' region of the gene encoding the sarcolemmal-associated proteins (SLAPs) defines a novel class of coiled-coil tail-anchored membrane proteins. J. Biol. Chem. 275:38474-38481 (2000).
		432556|SLMAP	2.7E-185	[Homo sapiens] [Endoplasmic reticulum; Cytoplasmic; Plasma membrane] Sarcolemmal-associated protein, a coiled-coil tail-anchored membrane protein that has alternatively spliced forms which are conserved between mammalian and avian species; expressed in sarcolemma of cardiac, soleus, and smooth muscle
				Wigle, J.T. et al. Molecular cloning, expression, and chromosomal assignment of sarcolemmal-associated proteins. A family of acidic amphipathic alpha-helical proteins associated with the membrane. J. Biol. Chem. 272:32384-32394 (1997).
				Wielowieyski, P.A. et al. (2000), *supra.*
		746477|Slap	1.5E-152	[Mus musculus] Sarcolemmal-associated protein, a coiled-coil tail-anchored membrane protein that has alternatively spliced forms which are conserved between mammalian and avian species
				Wielowieyski, P.A. et al. (2000), *supra.*
43	7515570CD1	g3790610	6.0E-92	[Cricetulus griseus] layilin
				Borowsky, M.L. and Hynes, R.O. Layilin, a novel talin-binding transmembrane protein homologous with C-type lectins, is localized in membrane ruffles. J. Cell Biol. 143:429-442 (1998).
44	7515680CD1	g6252444	8.4E-87	[Homo sapiens] endomucin-2
				Kinoshita, M. et al. Identification of human endomucin-1 and -2 as membrane-bound O-sialoglycoproteins with anti-adhesive activity, FEBS Lett. 499, 121-126 (2001)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		476331| LOC51705	6.6E-88	[Homo sapiens] Protein with high similarity to endomucin-1 (human LOC51169), which is a membrane attached O-sialoglycoprotein that may negatively regulate cell-matrix adhesion
		587741|Emcn	5.2E-40	[Mus musculus] Endomucin, a putative membrane-bound O-sialoglycoprotein, involved in cell adhesion
				Samulowitz, U. et al. Human endomucin: distribution pattern, expression on high endothelial venules, and decoration with the MECA-79 epitope. Am J Pathol 160, 1669-81. (2002).
45	7516698CD1	g11612079	1.1E-195	[Homo sapiens] DC-specific transmembrane protein
				Hartgers, F. C. et al. DC-STAMP, a novel multimembrane-spanning molecule preferentially expressed by dendritic cells, Eur. J. Immunol. 30, 3585-3590 (2000)
		700850| LOC81501	8.3E-197	[Homo sapiens] DC-specific transmembrane protein, a putative seven transmembrane spanning protein unrelated to other multimembrane-spanning receptors, expressed on activated but not resting blood dendritic cells, may function as a receptor or a cytoskeletal anchor
				Hartgers, F. C. et al. Genomic organization, chromosomal localization, and 5' upstream region of the human DC-STAMP gene. Immunogenetics 53, 145-9. (2001).
46	7517501CD1	g1915976	0.0	[Homo sapiens] CSF-1 receptor
				Andre, C. et al. Sequence analysis of two genomic regions containing the KIT and the FMS receptor tyrosine kinase genes, Genomics 39, 216-226 (1997)
		341990|CSF1R	0.0	[Homo sapiens][Protein kinase; Transferase; Receptor (signalling)][Plasma membrane] Colony stimulating factor 1 receptor, a tyrosine kinase receptor involved in the growth and differentiation of cells of the myeloid lineage; may be associated with invasive growth in breast cancer
				Stanley, E. R. et al. Biology and action of colony--stimulating factor-1. Mol Reprod Dev 46, 4-10 (1997).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		583877|Csf1r	4.3E-297	[Mus musculus][Protein kinase; Transferase; Receptor (signalling)][Plasma membrane] Colony stimulating factor 1 receptor, a tyrosine kinase receptor involved in the growth and differentiation of cells of the myeloid lineage; may be associated with invasive growth in cancer
				Gisselbrecht, S. et al. Frequent c-fms activation by proviral insertion in mouse myeloblastic leukaemias. Nature 329, 259-61 (1987).
47	7518576CD1	g2052056	1.0E-23	[Homo sapiens] SIRP-alpha1
				Kharitonenkov, A. et al. A family of proteins that inhibit signalling through tyrosine kinase receptors, Nature 386, 181-186 (1997)
		341116|PTPNS1	7.8E-25	[Homo sapiens][Plasma membrane] Protein tyrosine phosphatase non-receptor type substrate 1, a receptor-like protein that is a substrate for phosphotyrosine phosphatases and a regulatory protein in signal transduction pathways
				Yamao, T. et al. Mouse and human SHPS-1: molecular cloning of cDNAs and chromosomal localization of genes. Biochem Biophys Res Commun 231, 61-7 (1997).
				Oldenborg, P.-A. et al. Role of CD47 as a marker of self on red blood cells. Science 288,2051-2054 (2000).
		694548|Ptpns1	8.1E-24	[Rattus norvegicus][Receptor (signalling)][Plasma membrane] Protein tyrosine phosphatase non-receptor type substrate 1, a receptor-like protein that is a substrate for phosphotyrosine phosphatases and a regulatory protein in signal transduction pathways, has a role in cell adhesion signaling
				Fujioka, Y. et al. A novel membrane glycoprotein, SHPS-1, that binds the SH2-domain-containing protein tyrosine phosphatase SHP-2 in response to mitogens and cell adhesion. Mol Cell Biol 16, 6887-99 (1996).
48	7518626CD1	g7406952	7.7E-77	[Homo sapiens] 8D6 antigen
				Li, L. et al. Identification of a human follicular dendritic cell molecule that stimulates germinal center B cell growth, J. Exp. Med. 191; 1077-1084 (2000)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		476035|8D6A	6.0E-78	[Homo sapiens] 8D6 antigen, stimulates germinal center B cell growth and differentiation
				Zhang, X. et al. The distinct roles of T cell-derived cytokines and a novel follicular dendritic cell-signaling molecule 8D6 in germinal center-B cell differentiation. J Immunol 167, 49-56. (2001).
		608328| 425O18-1	1.6E-36	[Mus musculus][Small molecule-binding protein] Protein with high similarity to 8D6 antigen (human 8D6A), which stimulates germinal center B cell growth and differentiation, contains two low-density lipoprotein receptor class A domains
49	7515714CD1	g2052058	1.0E-58	[Homo sapiens] SIRP-beta1
				Kharitonenkov, A. et al. A family of proteins that inhibit signalling through tyrosine kinase receptors, Nature 386, 181-186 (1997)
		343192|SIRPB1	7.9E-60	[Homo sapiens] Signal regulatory protein beta 1, associates with phosphorylated DAP12 (TYROBP), leading to efficient cell-surface expression, recruitment of tyrosine kinase SYK, and tyrosine phosphorylation of ERK1 (MAPK3)/ERK2 (MAPK1) and cell activation
				Dietrich, J. et al. Cutting edge: signal-regulatory protein beta 1 is a DAP12-associated activating receptor expressed in myeloid cells. J Immunol 164, 9-12. (2000).
				Tomasello, E. et al. Association of signal-regulatory proteins beta with KARAP/DAP-12. Eur. J. Immunol. 30,2147-2156. (2000).
		585647|Ptpns1	4.3E-36	[Mus musculus][Receptor (signalling)] Protein tyrosine phosphatase non-receptor type substrate 1, a receptor-like protein that is a substrate for phosphotyrosine phosphatases and a regulatory protein in signal transduction pathways, has a role in cell adhesion signaling
				Comu, S. et al. The murine P84 neural adhesion molecule is SHPS-1, a member of the phosphatase-binding protein family. J Neurosci 17, 8702-10 (1997).

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	2847449CD1	102	Signal_cleavage: M1-T45	SPSCAN
			Cytosolic domain: K52-R102; Transmembrane domain: S29-L51; Non-cytosolic domain: M1-S28	TMHMMER
			Potential Phosphorylation Sites: S93 T18 T19 T60 Y26	MOTIFS
			Potential Glycosylation Sites: N16 N82	MOTIFS
2	7523642CD1	37	Signal_cleavage: M1-G29	SPSCAN
			Signal Peptide: M1-L19	HMMER
			Signal Peptide: M3-E31	HMMER
			Signal Peptide: M1-E31	HMMER
			Cytosolic domain: M1-F6; Transmembrane domain: L7-P26; Non-cytosolic domain: H27-F37	TMHMMER
			Secretin receptor signature PR00490: N13-M25	BLIMPS_PRINTS
			Potential Phosphorylation Sites: S28	MOTIFS
3	7521994CD1	379	Signal_cleavage: M1-S13	SPSCAN
			Signal Peptide: M1-Y15	HMMER
			Immunoglobulin: T185-H263, R93-S177	HMMER_SMART
			Immunoglobulin C-2 Type: L191-G252	HMMER_SMART
			Immunoglobulin domain: G193-A247, G101-V159	HMMER_PFAM
			Ig superfamily from SCOP: I181-L254	HMMER_INCY
			Cytosolic domain: R295-K379; Transmembrane domain: G272-F294; Non-cytosolic domain: M1-G271	TMHMMER
			OB BINDING MYELOID CELL SURFACE ANTIGEN CD33 PRECURSOR GP67 GLYCOPROTEIN PD015772: Q248-K379	BLAST_PRODOM
			CELL PRECURSOR GLYCOPROTEIN TRANSMEMBRANE SIGNAL IMMUNOGLOBULIN FOLD ADHESION ALTERNATIVE SPLICING PD005007: M1-V27 S91-Q173 A12-L155	BLAST_PRODOM
			MYELIN; SCHWANN; SIALOADHESIN; FORM; DM03744|P20138|1-142: L2-S13 Y17-V94	BLAST_DOMO
			Potential Phosphorylation Sites: S60 S87 S91 S178 S237 S344 T145 T244 T298 T368 Y72 Y372	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Glycosylation Sites: N56 N105 N154 N175 N184	MOTIFS
			Ig_Mhc: F243-H249	MOTIFS
4	7522289CD1	101	Low-density lipoprotein-receptor YWTD domain: Q26-H68	HMMER_SMART
			Low-density lipoprotein receptor repeat: R46-E87, G3-V44	HMMER_PFAM
			GLYCOPROTEIN PROTEIN RECEPTOR EGFLIKE DOMAIN LIPOPROTEIN PRECURSOR SIGNAL TRANSMEMBRANE RECEPTORRELATED PD149641: V5-S91	BLAST_PRODOM
			Potential Phosphorylation Sites: S67 S93 T8 T53	MOTIFS
			Potential Glycosylation Sites: N64	MOTIFS
5	7522336CD1	128	Signal_cleavage: M1-G65	SPSCAN
			Cytosolic domain: S63-A128; Transmembrane domain: L40-V62; Non-cytosolic domain: M1-I39	TMHMMER
			Potential Phosphorylation Sites: S14 S18 S19 S98 T26 T79 T93 T115 T120	MOTIFS
			Potential Glycosylation Sites: N77 N91 N96	MOTIFS
6	7522339CD1	121	Signal_cleavage: M1-S61	SPSCAN
			Cytosolic domain: S57-P121; Transmembrane domain: F34-L56; Non-cytosolic domain: M1-S33	TMHMMER
			Potential Phosphorylation Sites: S73 S78 S102 S116	MOTIFS
			Potential Glycosylation Sites: N106	MOTIFS
7	7522361CD1	183	Signal_cleavage: M1-T29	SPSCAN
			Signal Peptide: M1-A27	HMMER
			Cytosolic domains: M1-T15, S119-P183; Transmembrane domains: N16-L38, F96-L118; Non-cytosolic domain: E39-S95	TMHMMER
			Potential Phosphorylation Sites: S135 S140 S164 S178 T48	MOTIFS
			Potential Glycosylation Sites: N168	MOTIFS
8	7522368CD1	181	Signal_cleavage: M1-C64	SPSCAN
			C-type lectin (CTL) or carbohydrate-recognition domain (CRD): C64-Q180	HMMER_SMART
			Lectin C-type domain: E81-Y181	HMMER_PFAM
			Cytosolic domain: V62-Y181; Transmembrane domain: I39-L61; Non-cytosolic domain: M1-K38	TMHMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			C-type lectin domain IPB001304: W68-C92, W122-W134	BLIMPS_BLOCKS
			Type II antifreeze protein signature PR00356: L63-C75, C75-C92, L93-F110, W122-D138	BLIMPS_PRINTS
			C-TYPE LECTIN DM00035|I38700|87-211: L61-Q180	BLAST_DOMO
			C-TYPE LECTIN DM00035|P27811|87-211: C64-Q180	BLAST_DOMO
			C-TYPE LECTIN DM00035|P27812|87-211: C64-Q180	BLAST_DOMO
			C-TYPE LECTIN DM00035|P27471|87-211: C64-Q180	BLAST_DOMO
			Potential Phosphorylation Sites: S14 S18 S19 S84 S129 S140 T26 T135 Y91	MOTIFS
			Potential Glycosylation Sites: N126	MOTIFS
9	7522373CD1	85	Cytosolic domain: K73-V85; Transmembrane domain: I50-A72; Non-cytosolic domain: M1-G49	TMHMMER
			C5A-anaphylatoxin receptor signature PR00426: L55-S66	BLIMPS_PRINTS
			Potential Phosphorylation Sites: S14 S18 S19 T26	MOTIFS
10	7522381CD1	78	Signal_cleavage: M1-A65	SPSCAN
			Cytosolic domain: G62-V78; Transmembrane domain: I39-L61; Non-cytosolic domain: M1-K38	TMHMMER
			Potential Phosphorylation Sites: S14 S18 S19 T26	MOTIFS
11	7523596CD1	332	Signal_cleavage: M1-A29	SPSCAN
			Signal Peptide: M1-L19	HMMER
			Immunoglobulin: S37-A144	HMMER_SMART
			Immunoglobulin V-Type: D47-F128	HMMER_SMART
			V type Ig domains from SCOP: F31-L150	HMMER_INCY
			b7/CD80/CD86 Multiple Ig domain protein: V33-W248	HMMER_INCY
			Cytosolic domain: Q272-P332; Transmembrane domain: I249-W271; Non-cytosolic domain: M1-W248	TMHMMER
			BUTYROPHILIN PRECURSOR BT TRANSMEMBRANE GLYCOPROTEIN IMMUNOGLOBULIN FOLD SIGNAL PROTEIN BTF4 PD004895: Y127-I209	BLAST_PRODOM
			BUTYROPHILIN PRECURSOR SIGNAL IMMUNOGLOBULIN PROTEIN TRANSMEMBRANE GLYCOPROTEIN FOLD MYELINOLIGODENDROCYTE MYELIN PD000570: F31-C126	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			BUTYROPHILIN BTF4 PROTEIN BTF5 PUT B7 MOLECULE OF CD80CD86 FAMILY PD150098: M210-Q246	BLAST_PRODOM
			BUTYROPHILIN BTF4 PROTEIN BTF5 PUT B7 MOLECULE OF CD80CD86 FAMILY PD009489: R247-L290	BLAST_PRODOM
			IG; ANTIGEN; V-REGION-LIKE; B-G; DM02854|P18892|1-77: M3-V78	BLAST_DOMO
			IG; ANTIGEN; V-REGION-LIKE; B-G; DM02854|A47712|1-83: M3-E84	BLAST_DOMO
			Potential Phosphorylation Sites: S71 S121 S213 S236 S303 T117 T221 T297	MOTIFS
			Potential Glycosylation Sites: N115	MOTIFS
12	7523643CD1	533	Signal_cleavage: M1-G26	SPSCAN
			Signal Peptide: M1-A20	HMMER
			Signal Peptide: M1-T24	HMMER
			Signal Peptide: M1-G31	HMMER
			Repeats in polycystic kidney disease 1 (PKD1): S132-D225	HMMER_SMART
			PKD domain: V131-V220	HMMER_PFAM
			Cytosolic domain: Y488-V533; Transmembrane domain: L465-I487; Non-cytosolic domain: M1-G464	TMHMMER
			PKD domain proteins. PF00801: K25-P37, G177-V204, G432-T436	BLIMPS_PFAM
			PROTEIN PRECURSOR SIGNAL GLYCOPROTEIN TRANSMEMBRANE REPEAT MELANOCYTE PMEL PUTATIVE NMB PD144293: G324-V533	BLAST_PRODOM
			MELANOCYTE PROTEIN PMEL 17 PRECURSOR LINEAGESPECIFIC ANTIGEN GP100 MELANOMAASSOCIATED ME20 ME20M/ME20S ME20M/ME20S 95 KD MELANOCYTESPECIFIC SECRETED GLY PD144289: A202-L345 S217-T323	BLAST_PRODOM
			PROTEIN PRECURSOR SIGNAL GLYCOPROTEIN TRANSMEMBRANE REPEAT MELANOCYTE PMEL PUTATIVE NMB PD008898: G26-V131	BLAST_PRODOM
			NMB; PMEL; DM04736|P40967|388-667: T263-V533	BLAST_DOMO
			NMB; PMEL; DM04737|P40967|1-272: M1-G32 G16-A187	BLAST_DOMO
			NMB; PMEL; DM04736|S53871|377-626: A288-Q532	BLAST_DOMO
			NMB; PMEL; DM04737|S53871|1-276: G26-T191	BLAST_DOMO
			Potential Phosphorylation Sites: S58 S108 S442 T88 T122 T193 T291 T299 T335 T396	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Glycosylation Sites: N235 N440	MOTIFS
13	7523769CD1	311	Signal_cleavage: M1-A29	SPSCAN
			Signal Peptide: M1-L19	HMMER
			Signal Peptide: M3-F31	HMMER
			Immunoglobulin: S37-A144	HMMER_SMART
			Immunoglobulin V-Type: D47-F128	HMMER_SMART
			V type Ig domains from SCOP: F31-R150	HMMER_INCY
			b7/CD80/CD86 multiple Ig domain p: V33-A249	HMMER_INCY
			BUTYROPHILIN PRECURSOR SIGNAL IMMUNOGLOBULIN PROTEIN TRANSMEMBRANE GLYCOPROTEIN FOLD MYELINOLIGODENDROCYTE MYELIN PD000570: F31-C126	BLAST_PRODOM
			IG; ANTIGEN; V-REGION-LIKE; B-G; DM02854|P18892|1-77: M3-V78	BLAST_DOMO
			IG; ANTIGEN; V-REGION-LIKE; B-G; DM02854|A47712|1-83: M3-E84	BLAST_DOMO
			Potential Phosphorylation Sites: S71 S121 S195 S269 S304 T117	MOTIFS
			Potential Glycosylation Sites: N115	MOTIFS
14	7523785CD1	419	Signal Peptide: M1-L19	HMMER
			signal_cleavage: M1-A29	SPSCAN
			SPRY domain: S298-A419	HMMER_PFAM
			Immunoglobulin: S37-A144	HMMER_SMART
			Immunoglobulin V-Type: D47-F128	HMMER_SMART
			Domain in SPla and the RYanodine Receptor: S298-A419	HMMER_SMART
			V type Ig domains from SCOP (antibody variable domain): F31-S150	HMMER_INCY
			Ig superfamily from SCOP: V33-F147	HMMER_INCY
			b7/CD80/CD86 multiple Ig domain proteins: V33-A234	HMMER_INCY
			Non-cytosolic domain: M1-W154 Transmembrane domain: I155-W177 Cytosolic domain: Q178-A419	TMHMMER
			Butyrophylin C-terminal DUF signature PR01407: K244-V261, V261-P278, R283-V307, W313-G326, L357-N381, I388-I406	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Domain in SPla and the RYanodine Receptor PF00622: E135-F148, E282-W303, V366-F379	BLIMPS_PFAM
			BUTYROPHILIN SIGNAL PRECURSOR GLYCOPROTEIN IMMUNOGLOBULIN TRANSMEMBRANE DOMAIN MYELIN-OLIGODENDROCYTE B-G MHC PD000570: F31-C126	BLAST_PRODOM
			RECEPTOR CHANNEL RYANODINE CALCIUM MUSCLE IONIC REPEAT SKELETAL TRANSMEMBRANE CALCIUM-BINDING PHOSPHORYLATION BUTYROPHILIN PD001178: S298-P402	BLAST_PRODOM
			ZINC-FINGER BUTYROPHILIN RING NUCLEAR COIL COILED METAL-BINDING MIDLINE ANTIGEN RING STONUS TOXIN PUTATIVE TRANSCRIPTION FACTOR XPRF PD002421: L250-P418	BLAST_PRODOM
			ZINC-FINGER BUTYROPHILIN NUCLEAR ANTIGEN DNA-BINDING RET RIBONUCLEOPROTEIN RNA-BINDING PD002445: R197-F296	BLAST_PRODOM
			RFP TRANSFORMING PROTEIN DM01944|P18892|355-477: S298-P418 DM01944|A43906|483-608: S298-P402 DM01944|Q02084|497-622: S298-P402 DM01944|P14373|368-492: S298-P418	BLAST_DOMO
			Potential Phosphorylation Sites: S71, S121, S209, S227, S298, S395, T117, T203, T349, T390	MOTIFS
			Potential Glycosylation Sites: N115	MOTIFS
15	7523836CD1	539	Signal Peptide: M1-C17, M1-L19, M1-S20, M1-T22, M1-L24, M1-A25, M1-T27	HMMER
			signal_cleavage: M1-A25	SPSCAN
			Non-cytosolic domain: M1-A505 Transmembrane domain: V506-Y528 Cytosolic domain: S529-Q539	TMHMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			RECEPTOR BETAGLYCAN III TYPE TGFR-3 PROTEOGLYCAN PRECURSOR SIGNAL GLYCOPROTEIN TRANSMEMBRANE ENDOGLIN CELL PD006478: L8-P142, L8-K347	BLAST_PRODOM
			RECEPTOR TRANSMEMBRANE SIGNAL PRECURSOR GLYCOPROTEIN BETAGLYCAN III TYPE CELL TGFR-3 ENDOGLIN CELL PD009789: G500-E533	BLAST_PRODOM
			ENDOGLIN PRECURSOR CELL ADHESION GLYCOPROTEIN TRANSMEMBRANE SIGNAL ANTIGEN CD105 DISEASE PD152706: A327-Q476	BLAST_PRODOM
			ENDOGLIN DM08280|S50831|1-658: M1-Q476, S449-R532 DM08280|P37176|1-652: M1-Q476, S464-R532	BLAST_DOMO
			Potential Phosphorylation Sites: S23, S76, S158, S278, S313, T126, T235, T273, T388, T497	MOTIFS
			Potential Glycosylation Sites: N88, N102, N121, N134, N307	MOTIFS
			Cell attachment sequence: R399-D401	MOTIFS
16	7523879CD1	558	Signal Peptide: M1-P32	HMMER
			signal_cleavage: M1-P32	SPSCAN
			Fibronectin type III domain: A242-S328, L339-G423	HMMER_PFAM
			Non-cytosolic domain: M1-K438 Transmembrane domain: V439-S461 Cytosolic domain: G462-A558	TMHMMER
			CLASS I CYTOKINE RECEPTOR CRL1 PD043640: T30-P363, P180-L549 PD145804: M1-R29	BLAST_PRODOM
			Potential Phosphorylation Sites: S70, S75, S100, S140, S158, S187, S213, S405, S461, S533, T98, T226, T398, T434, T542	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Glycosylation Sites: N51, N76, N224, N233, N296, N304, N389, N485	MOTIFS
			Growth factor and cytokines receptors family signature 1: C41-W54	MOTIFS
17	7523880CD1	512	Signal Peptide: M1-P18, M1-A22	HMMER
			signal_cleavage: M1-A22	SPSCAN
			Repeats in polycystic kidney disease 1 (PKD1) and other proteins: S202-G326	HMMER_SMART
			Non-cytosolic domain: M1-M436 Transmembrane domain: A437-Y459 Cytosolic domain: K460-S512	TMHMMER
			PKD domain proteins PF00801: G238-L265	BLIMPS_PFAM
			TRANSMEMBRANE GLYCOPROTEIN PRECURSOR SIGNAL MELANOCYTE REPEAT PMEL ANTIGEN OSTEOACTIVIN SECRETED PUTATIVE NMB PD008898:W54-K74, G65-V192	BLAST_PRODOM
			PROTEIN PRECURSOR SIGNAL GLYCOPROTEIN TRANSMEMBRANE REPEAT MELANOCYTE PMEL PUTATIVE NMB PD144293: S290-Y459	BLAST_PRODOM
			TRANSMEMBRANE GLYCOPROTEIN SIGNAL PRECURSOR NMB OSTEOACTIVIN QNR-71 PD144295: S432-Q506	BLAST_PRODOM
			TRANSMEMBRANE GLYCOPROTEIN OSTEOACTIVIN NMB PRECURSOR SIGNAL PD179435: M1-D53	BLAST_PRODOM
			NMB; PMEL DM04736|I38065|296-560: N248-V511 DM04737|I38065|1-294: M1-A80, K70-T247 DM04737|S53871|1-276: W54-S79, V107-S246 DM04737|P40967|1-272: W54-K112, V107-S246	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Phosphorylation Sites: S48, S72, S94, S143, S202, S348, S382, S424, T266, T322, T409	MOTIFS
			Potential Glycosylation Sites: N86, N98, N152, N201, N227, N248, N252, N258, N264, N399, N407	MOTIFS
			Cell attachment sequence: R64-D66	MOTIFS
18	7523812CD1	168	Signal Peptide: M1-A16, M1-G19, M1-L22	HMMER
			signal_cleavage: M1-A16	SPSCAN
			Ephrin receptor ligand binding domain: E18-G148	HMMER_PFAM
			RECEPTOR KINASE TYROSINE PROTEIN ATP-BINDING TRANSFERASE EPHRIN PRECURSOR PHOSPHORYLATION TRANSMEMBRANE GLYCOPROTEIN PD001495: E18-K139	BLAST_PRODOM
			RECEPTOR TYROSINE KINASE CLASS V DM00501|P29323|22-375:L21-K139 DM00501|A56599|30-383:L21-K139 DM00501|P28693|30-383:L21-K139 DM00501|P54758|35-381:V20-K135	BLAST_DOMO
			Potential Phosphorylation Sites: S45, S95, S163, T24, T54, T69, T108, T138	MOTIFS
19	7524026CD1	291	Signal Peptide: M1-C24, M1-S20, M1-L27, M1-S22	HMMER
			signal_cleavage: M1-S20	SPSCAN
			TGF-beta propeptide: C24-H255	HMMER_PFAM
			Cytosolic domain: M1-C4 Transmembrane domain: V5-L27 Non-cytosolic domain: D28-I291	TMHMMER
			Transforming growth factor beta precursor signature PR01423: A8-L27, E62-T74, V161-K175, T181-G196, W198-H212	BLIMPS_PRINTS
			Transforming growth factor beta 2 precursor signature PR01425: H2-T23, L50-S69, D76-Y98, N116-R131, F152-E163, L172-Y184, P233-R249	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Transforming growth factor beta 3 precursor signature PR01426: Y3-S22, D76-R93, V151-E163, K175-A194, L211-H225	BLIMPS_PRINTS
			GLYCOPROTEIN FACTOR GROWTH PRECURSOR BONE SIGNAL MORPHOGENETIC CYTOKINE TGF-BETA CARTILAGE PD000853: L19-K173, S52-P258	BLAST_PRODOM
			TGF-BETA FAMILY DM00245|A34005|49-414:K49-L253 DM00245|P30371|49-412:K49-L253 DM00245|P16047|51-412:L48-F250 DM00245|P16176|50-382:K49-T230	BLAST_DOMO
			Potential Phosphorylation Sites: S25, S52, S73, S83, S94, S135, T181, T192, Y98, Y99	MOTIFS
			Potential Glycosylation Sites: N72, N140, N241	MOTIFS
20	7524357CD1	490	Signal Peptide: M1-V27, M1-P23	HMMER
			signal_cleavage: M1-N26	SPSCAN
			Cytosolic domain: M1-N4 Transmembrane domain: F5-V27 Non-cytosolic domain: T28-T490	TMHMMER
			BRUSH BORDER 61.9KD PROTEIN PRECURSOR SIGNAL PD144534: L8-N292, S52-K367, W298-C486	BLAST_PRODOM
			Potential Phosphorylation Sites: S71, S121, S150, S216, S258, S294, S316, S405, T28, T41, T274	MOTIFS
			Potential Glycosylation Sites: N26, N237, N292	MOTIFS
21	7524808CD1	407	Signal Peptide: M1-G32, P12-G32	HMMER
			Signal cleavage: M1-G32	SPSCAN
			SPRY domain: S263-L383	HMMER_PFAM
			Domain in SPla and the RYanodine Receptor: S263-L383	HMMER_SMART

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Cytosolic domains: S35-T120, K171-L407 Transmembrane domains: L15-G34, V121-L143, W148-I170 Non-cytosolic domains: M1-S14, T144-P147	TMHMMER
			Butyrophylin C-terminal DUF signature PR01407: H209-L226, L226-P243, R248-V272, W278-G291, L321-N345, I352-L370	BLIMPS_PRINTS
			Domain in SPla and the RYanodine Receptor PF00622: E247-W268, V330-F343	BLIMPS_PFAM
			ZINC-FINGER BUTYROPHILIN RING NUCLEAR COIL COILED METAL-BINDING MIDLINE ANTIGEN PD002421: L215-F368	BLAST_PRODOM
			ZINC-FINGER BUTYROPHILIN NUCLEAR ANTIGEN DNA-BINDING RET RIBONUCLEOPROTEIN RNA-BINDING PD002445: I170-F261	BLAST_PRODOM
			BUTYROPHILIN MEMBER GLYCOPROTEIN SIGNAL SUBFAMILY DOMAIN PRECURSOR BUTYROPHILIN-LIKE BT TRANSMEMBRANE PD004895: V30-I96	BLAST_PRODOM
			BUTYROPHILIN PRECURSOR DOMAIN BT SIGNAL GLYCOPROTEIN IMMUNOGLOBULIN TRANSMEMBRANE BK14H9.2 PD021641: I97-A194	BLAST_PRODOM
			RFP TRANSFORMING PROTEIN DM01944|P14373|368-492: S263-P381 DM01944|P18892|355-477: S263-G385 DM01944|P19474|339-465: S263-P381 DM01944|Q02084|497-622: S263-S375	BLAST_DOMO
			Potential Phosphorylation Sites: S81, S232, S263, S315	MOTIFS
			Potential Glycosylation Sites: N104, N110	MOTIFS
22	7522161CD1	252	Zinc finger, C4 type (two domains): E96-E171	HMMER_PFAM
			C4 zinc finger in nuclear hormone receptor: N95-E166	HMMER_SMART
			C4-type steroid receptor zinc finger IPB001628: C98-V126, Y132-V168	BLIMPS_BLOCKS
			Nuclear hormones receptors DNA-binding region signature: S29-T144	PROFILESCAN
			C4-type steroid receptor zinc finger signature PR00047: C98-S114, S114-G129, R147-L155, L155-M163	BLIMPS_PRINTS
			Vitamin D receptor signature PR00350: C98-S114, C115-C134	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			RECEPTOR NUCLEAR TRANSCRIPTION REGULATION DNA-BINDING ZINC-FINGER HORMONE FAMILY ORPHAN ALTERNATIVE PD000035: L97-M163	BLAST_PRODOM
			RECEPTOR NUCLEAR PROTEIN TRANSCRIPTION REGULATION DNABINDING ZINCFINGER ORPHAN UBIQUITOUSLY EXPRESSED PD010811: Q205-T236	BLAST_PRODOM
			RECEPTOR LXR-ALPHA NUCLEAR REGULATION ZINC-FINGER LIVER OXYSTEROLS TRANSCRIPTION PD029416: M1-E96	BLAST_PRODOM
			NUCLEAR RECEPTOR TRANSCRIPTION REGULATION DNA BINDING PROTEIN ZINC FINGER ORPHAN LXR ALPHA PD032282: R164-P204	BLAST_PRODOM
			NUCLEAR HORMONES RECEPTORS DNA-BINDING REGION DM00047|P55055|77-389:P88-P208 Q205-T236 DM00047|P49881|196-515:K86-V218 L203-P252 DM00047|I38975|68-375: T68-T236 DM00047|I49021|68-374: P88-Q222	BLAST_DOMO
			Potential Phosphorylation Sites: S16, S228, S230, T68	MOTIFS
			Potential Glycosylation Sites: N241	MOTIFS
			Type-1 copper (blue) proteins signature: G129-M142	MOTIFS
			Nuclear hormones receptors DNA-binding region signature: C98-R124	MOTIFS
			Zinc finger, C2H2 type, domain: C115-H135	MOTIFS
23	7523999CD1	1473	Signal Peptide: M1-G21	HMMER
			Signal cleavage: M1-G21	SPSCAN
			Fibronectin type III domain: P526-S611, L642-S728, P740-S830	HMMER_PFAM
			Immunoglobulin domain:G45-A112, G147-G205, E239-A295, G332-A397, D436-A494	HMMER_PFAM
			Fibronectin type 3 domain: P526-S608, L642-G725, P740-G827	HMMER_SMART
			Immunoglobulin:P37-A129, P139-F222, P231-R311, P324-T413, P428-T510	HMMER_SMART
			Immunoglobulin C-2 Type:S43-G117, A145-G210, L237-G300, A330-G402, A434-G499	HMMER_SMART
			Immunoglobulin V-Type:A241-A295, T334-A397, T438-A494	HMMER_SMART
			I type Ig domains from SCOP:P30-F135, Q137-L228, R229-P317, P318-R419, P420-S517	HMMER_INCY
			V type Ig domains from SCOP (anti:D133-L228, P318-P422, I423-I516	HMMER_INCY
			Ig superfamily from SCOP:I33-L130, F135-P225, F227-A316, F320-I423, I424-T501	HMMER_INCY

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Cytosolic domain: R887-L1473 Transmembrane domain: A864-W886 Non-cytosolic domain: M1-P863	TMHMMER
			ROUNDABOUT REPEAT TRANSMEMBRANE RIG-1 RECEPTOR ROBO1 DUTT1 CG14348 CG5423 PD143949:K967-N1048 P1020-G1258 V828-Q1066 P1198-P1325	BLAST_PRODOM
			C33F10.5A SIMILAR TO FIBRONECTIN TYPE III PD073047: P225-D414	BLAST_PRODOM
			TRANSMEMBRANE RECEPTOR ROUNDABOUT ROBO1 DUTT1 PD112973: F1261-G1459	BLAST_PRODOM
			BASIC FIBROBLAST GROWTH FACTOR RECEPTOR 1 DM01287|A39752|1-814:G147-V204 V329-E511 E240-P420	BLAST_DOMO
			BASIC FIBROBLAST GROWTH FACTOR RECEPTOR 1 DM01287|P16092|1-821:E240-G300 R313-E511 S1062-Q1099 E240-C443, T164-G210	BLAST_DOMO
			BASIC FIBROBLAST GROWTH FACTOR RECEPTOR 1 DM01287|P18460|1-805:D141-C203 V329-E511	BLAST_DOMO
			BASIC FIBROBLAST GROWTH FACTOR RECEPTOR 1 DM01287|A60350|1-800:E240-C293 V329-E511 P959-S1027	BLAST_DOMO
			Potential Phosphorylation Sites: S43, S80, S101, S182, S285, S497, S851, S979, S1011, S1040, S1062, S1160, S1214, S1247, S1264, S1286, S1351, S1359, S1360, S1366, S1387, S1431, T73, T193, T220, T255, T277, T309, T417, T470, T585, T697, T732, T735, T984,	MOTIFS
			T1089, T1257, T1292, T1374, T1389, Y108, Y166, Y201, Y490	
			Potential Glycosylation Sites: N123, N430, N756, N786, N793, N849, N911, N940, N990, N1061, N1069, N1278, N1451	MOTIFS
			Cell attachment sequence: R903-D905, R954-D956	MOTIFS
24	7524024CD1	778	Signal Peptide: M1-G21, M1-E19, M1-Q23, M1-L22	HMMER
			Immunoglobulin domain:G62-A129, G163-A229, D264-V316, G349-A400, G433-A501	HMMER_PFAM
			Immunoglobulin:P54-L144, G155-Q247, P341-N417, T425-Q517	HMMER_SMART
			Immunoglobulin C-2 Type:V60-I134, R161-I234, D347-V405	HMMER_SMART
			I type Ig domains from SCOP: G334-S423	HMMER_INCY
			Ig superfamily from SCOP:F50-D149, V151-I246, V252-T323, M337-P420, I421-Q517	HMMER_INCY

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Cytosolic domain: C559-V778 Transmembrane domain: V536-F558 Non-cytosolic domain: M1-A535	TMHMMER
			EG:163A10.1 DOMAIN IRREGULAR RST PRECURSOR IRREC SIGNAL ADHESION CELL KIRRE PD124347:F50-V256 V261-E315	BLAST_PRODOM
			IMMUNOGLOBULIN DM00001|Q08180|31-126: S51-T142	BLAST_DOMO
			Potential Phosphorylation Sites: S159, S207, S215, S272, S373, S387, S454, S465, S474, S507, S563, S572, S588, S702, S715, S721, S734, T230, T301, T384, T597, T642, T725, Y48, Y307, Y396	MOTIFS
			Potential Glycosylation Sites: N167, N253, N324, N498	MOTIFS
			Leucine zipper pattern: L8-L29	MOTIFS
25	7522455CD1	279	Signal cleavage: M1-G41	SPSCAN
			Zinc finger, C4 type (two domains): D111-E186	HMMER_PFAM
			C4 zinc finger in nuclear hormone recepto: F110-D181	HMMER_SMART
			C4-type steroid receptor zinc finger IPB001628: C113-I141, Y147-V183	BLIMPS_BLOCKS
			Nuclear hormones receptors DNA-binding region signature: Q95-H160	PROFILESCAN
			C4-type steroid receptor zinc finger signature PR00047: C113-T129, T129-N144, R162-L170, L170-M178	BLIMPS_PRINTS
			Steroid hormone receptor signature PR00398: I174-Q184	BLIMPS_PRINTS
			RECEPTOR NUCLEAR TRANSCRIPTION REGULATION DNA-BINDING ZINC-FINGER HORMONE FAMILY ORPHAN ALTERNATIVE PD000035: L112-M178	BLAST_PRODOM
			RECEPTOR NUCLEAR ZINC-FINGER DNA-BINDING REGULATION ORPHAN TRANSCRIPTION STEROID TR2 PD005785: K179-K262	BLAST_PRODOM
			RECEPTOR NUCLEAR DNA-BINDING ORPHAN REGULATION ZINC-FINGER TRANSCRIPTION TR2 STEROID PD006642: M1-D111	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			NUCLEAR HORMONES RECEPTORS DNA-BINDING REGION DM00047|B36738|103-518: D103-K261 DM00047|I80177|107-511: Q104-L276 DM00047|P49116|106-511: Q104-L276	BLAST_DOMO
			DNA; TR2-11; DM05510|B36738|1-101: M1-P102	BLAST_DOMO
			Potential Phosphorylation Sites: S121, S140, S148, S229, S259, T3, T33, T47, T72, T85, T204, T225	MOTIFS
			Nuclear hormones receptors DNA-binding region signature: C113-R139	MOTIFS
			Serine proteases, subtilase family, aspartic acid active site: T233-N243	MOTIFS
26	7524494CD1	1360	Signal Peptide: M1-G24	HMMER
			Signal cleavage: M1-G24	SPSCAN
			Plexin repeat: G526-P568	HMMER_PFAM
			Sema domain: M58-D507	HMMER_PFAM
			IPT/TIG domain:P569-D671, P684-R767, P770-L860	HMMER_PFAM
			Protein kinase domain: T1042-G1310	HMMER_PFAM
			ig-like, plexins, transcription factors:P568-M682, E683-R767, D769-L873	HMMER_SMART
			domain found in Plexins, Semaphorins and Integrins: G526-P568	HMMER_SMART
			Serine/Threonine protein kinases, catalytic domain: T1042-V1305	HMMER_SMART
			Tyrosine kinase, catalytic domain: T1042-V1301	HMMER_SMART
			Cytosolic domain: R943-T1360 Transmembrane domain: G920-W942 Non-cytosolic domain: M1-L919	TMHMMER
			Receptor tyrosine kinase class V IPB001426: Q1064-P1117, G1149-A1170, A1171-E1197, H1206-W1238, R1264-H1312, Q554-L601, E1239-G1263	BLIMPS_BLOCKS
			Receptor tyrosine kinase class III IPB001824: Q707-P739, E1081-S1135, D1144-K1183, H1204-P1246, P1246-L1297	BLIMPS_BLOCKS
			Receptor tyrosine kinase class II IPB002011: M1095-N1139, F1164-A1215, P1253-L1297	BLIMPS_BLOCKS
			Protein kinases signatures and profile: D1144-E1197	PROFILESCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Receptor tyrosine kinase class II signature: R1172-T1221	PROFILESCAN
			Tyrosine kinase catalytic domain signature PR00109: L1121-R1134, Y1158-L1176, L1209-L1219, S1228-H1250, C1272-F1294	BLIMPS_PRINTS
			RECEPTOR KINASE TYROSINE-PROTEIN TRANSFERASE ATP-BINDING PLEXIN PRECURSOR FACTOR GROWTH PD003981:C567-Q835 L884-Q946	BLAST_PRODOM
			SEMAPHORIN SIGNAL PRECURSOR GLYCOPROTEIN DEVELOPMENTAL NEUROGENESIS FAMILY MULTIGENE DOMAIN PD001844: D64-V474	BLAST_PRODOM
			RECEPTOR P185-RON ATP-BINDING MSP TRANSFERASE PHOSPHORYLATION ANTIGEN TYROSINE-PROTEIN PRECURSOR KINASE PD145274: G912-F1000	BLAST_PRODOM
			KINASE RECEPTOR PROTEIN TYROSINE CMETRELATED TRANSFORMING SEA PRECURSOR SIGNAL PD150595: V486-H566	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM00004|I48751|1061-1310: S1044-F1294 DM00004|Q04912|1083-1333: H1043-F1294	BLAST_DOMO
			do KINASE; TYROSINE; HEPATOCYTE; ATP; DM03653|I48751|19-535: D26-C533 DM03653|Q04912|17-533: L17-G534	BLAST_DOMO
			Potential Phosphorylation Sites: S201, S212, S217, S268, S284, S338, S421, S434, S490, S584, S623, S680, S802, S812, S843, S973, S999, S1001, S1003, S1016, S1044, T187, T286, T352, T376, T454, T713, T1042, T1141, T1181, T1221, T1225, T1226, T1327, Y44,	MOTIFS
			Y1198	
			Potential Glycosylation Sites: N419, N458, N488, N654, N720, N841	MOTIFS
			Cell attachment sequence: R845-D847	MOTIFS
			Protein kinases ATP-binding region signature: I1048-K1074	MOTIFS
			Tyrosine protein kinases specific active-site signature: F1164-L1176	MOTIFS
27	7524965CD1	74	Signal Peptide: M1-A38, M1-A34, S14-A35	HMMER
			Signal cleavage: M1-A38	SPSCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Cytosolic domain: M1-A20 Transmembrane domain: L21-L43 Non-cytosolic domain: L44-S74	TMHMMER
			PROTEASE SERINE TRANSMEMBRANE 3.4.21.- HYDROLASE HEPSIN SIGNAL-ANCHOR PD021735: K4-P53	BLAST_PRODOM
			Potential Phosphorylation Sites: S46, Y52	MOTIFS
			Potential Glycosylation Sites: N72	MOTIFS
28	7525018CD1	694	Signal Peptide: M1-A23, M1-Q20, M1-C24, M1-T26, M1-S18	HMMER
			Signal cleavage: M1-A23	SPSCAN
			Fibronectin type III domain: A446-S532	HMMER_PFAM
			Fibronectin type 3 domain: A446-G529	HMMER_SMART
			Cytosolic domain: G569-P694 Transmembrane domain: W546-L568 Non-cytosolic domain: M1-D545	TMHMMER
			Growth factor and cytokine receptor family 2 IPB002465: Q218-P228	BLIMPS_BLOCKS
			RECEPTOR BETA INTERLEUKIN-12 TRANSMEMBRANE COMPONENT CHAIN IL-12R-BETA1 REPEAT PRECURSOR SIGNAL PD032097: L527-E653	BLAST_PRODOM
			RECEPTOR PRECURSOR SIGNAL BETA INTERLEUKIN12 BETA1 CHAIN IL12R TRANSMEMBRANE GLYCOPROTEIN PD042763: G84-A446, L16-L197	BLAST_PRODOM
			IL12; GROWTH; COMPONENT; CYTOKINE; DM08716|I37892|206-397: Q206-T398 DM08716|P42701|206-397: Q206-T398 DM08720|I37892|513-662: G513-R657 DM08720|P42701|513-661: G513-L662	BLAST_DOMO
			Potential Phosphorylation Sites: S57, S83, S136, S453, S588, S612, S651, S665, T123, T257, T263, T281, T338	MOTIFS
			Potential Glycosylation Sites: N121, N329, N346, N352, N442, N456	MOTIFS
			N-6 Adenine-specific DNA methylases signature: C30-Y36	MOTIFS
			Growth factor and cytokines receptors family signature 2: S220-S226	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
29	7516620CD1	1719	Signal Peptide: M1-G19, M1-A25, M1-A17, M1-G19	HMMER
			Signal cleavage: M1-G19	SPSCAN
			Plexin repeat: T490-V540, N637-P684, K785-S838	HMMER_PFAM
			Sema domain: L33-D471	HMMER_PFAM
			IPT/TIG domain:P840-V933, P935-T1020, P1023-Y1122, P1125-S1211	HMMER_PFAM
			ig-like, plexins, transcription factors: H839-V933, T934-T1020, D1022-Y1122, D1124-P1219	HMMER_SMART
			Domain found in Plexins, Semaphorins and Integrins: T490-V540, N637-P684, K785-S838	HMMER_SMART
			Cytosolic domain: K1243-W1719 Transmembrane domain: A1220-Y1242 Non-cytosolic domain: M1-P1219	TMHMMER
			PLEXIN RECEPTOR PRECURSOR SIGNAL SEMAPHORIN TRANSMEMBRANE FIS HEMBA1005576 FOR MUSCULUS PD008852:P455-L486 T1479-N1718, A1237-S1647, C647-R676	BLAST_PRODOM
			RECEPTOR KINASE TYROSINE-PROTEIN TRANSFERASE ATP-BINDING PLEXIN PRECURSOR FACTOR GROWTH PD003981:C837-E927 D1022-P1193 V876-S1187	BLAST_PRODOM
			RECEPTOR KINASE TRANSFERASE TYROSINE-PROTEIN ATP-BINDING PRECURSOR PLEXIN SIGNAL GROWTH FACTOR PD003973: R352-H474	BLAST_PRODOM
			PLEXIN RECEPTOR PRECURSOR SIGNAL TRANSMEMBRANE PLEXIN-B3 PLEXA SEMAPHORIN SEX A3 PD010132: P546-H819	BLAST_PRODOM
			KINASE; TYROSINE; ATP; GROWTH; DM01368|P51805|796-899: C796-E900 DM01368|P51805|901-989: S901-P990 DM02937|P51805|991-1085: L991-L1086	BLAST_DOMO
			KINASE; TYROSINE; HEPATOCYTE; ATP; DM03653|P08581|14-526: D30-A498	BLAST_DOMO
			Potential Phosphorylation Sites: S82, S182, S183, S188, S274, S283, S436, S479, S598, S671, S778, S831, S958, S1158, S1357, S1374, S1410, S1462, S1521, S1596, S1606, S1610, S1707, T169, T248, T252, T259, T421, T614, T664, T703, T759, T924, T969, T1026, T1050, T1110, T1175, T1189, T1246, T1252, T1549, T1550, T1714, Y930, Y1515	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Glycosylation Sites: N59, N548, N637, N738, N746, N1009, N1036, N1073, N1115, N1162	MOTIFS
			ATP/GTP-binding site motif A (P-loop): G168-S175	MOTIFS
30	7525149CD1	672	Signal Peptide: M1-A21, M1-A16, M1-P23, M1-D30, M1-V28, M1-S26, M1-S19	HMMER
			Signal cleavage: M1-A21	SPSCAN
			Immunoglobulin domain: A57-V118, G156-A229	HMMER_PFAM
			Protein kinase domain: L366-R644	HMMER_PFAM
			Immunoglobulin:K49-V134, P148-L245	HMMER_SMART
			Immunoglobulin C-2 Type: P55-G123, V154-G234	HMMER_SMART
			Serine/Threonine protein kinases, catalytic domain: L366-L646	HMMER_SMART
			Tyrosine kinase, catalytic domain: L366-L642	HMMER_SMART
			I type Ig domains from SCOP: A38-R140	HMMER_INCY
			Ig superfamily from SCOP: E45-R140, Q144-R250	HMMER_INCY
			Cytosolic domain: C283-R672 Transmembrane domain: Y260-L282 Non-cytosolic domain: M1-D259	TMHMMER
			Receptor tyrosine kinase class V IPB001426: I492-A513, A514-D540, N547-W579, E580-G604, H605-E653, K392-L445	BLIMPS_BLOCKS
			Receptor tyrosine kinase class III IPB001824: K207-L245, T409-A463, D487-K526, T545-S587, S587-L638	BLIMPS_BLOCKS
			Receptor tyrosine kinase class II IPB002011: M423-P467, C507-E558, V594-L638, G90-N96	BLIMPS_BLOCKS
			Protein kinases signatures and profile: L488-N538	PROFILESCAN
			Receptor tyrosine kinase class II signature: R515-D562	PROFILESCAN
			Tyrosine kinase catalytic domain signature PR00109: V449-R462, Y501-V519, L550-L560, S569-G591, C613-F635	BLIMPS_PRINTS
			KINASE TRANSFERASE ATP-BINDING SERINE/THREONINE-PROTEIN TYROSINE-PROTEIN RECEPTOR 2.7.1.- PHOSPHORYLATION PRECURSOR PD000001:L368-L461 E481-I536 K553-K602	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			RECEPTOR KINASE GROWTH ATP-BINDING FACTOR TYROSINE-PROTEIN TRANSFERASE FIBROBLAST PRECURSOR SIGNAL PD001520: I307-G369	BLAST_PRODOM
			FIBROBLAST GROWTH FACTOR RECEPTOR 2 PRECURSOR PD009263: V155-G190	BLAST_PRODOM
			RECEPTOR ATP-BINDING KINASE FACTOR TRANSFERASE FIBROBLAST GROWTH FGFR-4 TYROSINE-PROTEIN PHOSPHORYLATION PD030900: G230-R310	BLAST_PRODOM
			BASIC FIBROBLAST GROWTH FACTOR RECEPTOR 1 DM01287|P16092|1-821: A38-R672 DM01287|P18460|1-805: E29-D670 DM01287|Q01742|1-654: M1-E652 DM01287|Q03364|1-812: A38-D670	BLAST_DOMO
			Potential Phosphorylation Sites: S26, S109, S338, S349, S418, S504, S630, T31, T42, T44, T59, T73, T205, T218, T288, T289, T304, T333, T409, T484, T546, T634, T649, T650, Y193, Y213	MOTIFS
			Potential Glycosylation Sites: N113, N126, N150, N182, N203, N216, N287, N612	MOTIFS
			Protein kinases ATP-binding region signature: L372-K402	MOTIFS
			Tyrosine protein kinases specific active-site signature: C507-V519	MOTIFS
31	7513047CD1	112	signal_cleavage: M1-A42	SPSCAN
			Signal Peptide: M1-A42, V19-A42, G20-A42, P22-A42	HMMER
			MRNA ENCODING RAMP2 PRECURSOR SIGNAL PD081157: M1-T48	BLAST_PRODOM
			Potential Phosphorylation Sites: S3, S84	MOTIFS
32	7513056CD1	65	signal_cleavage: M1-L24	SPSCAN
			PROTEIN NUCLEAR FACTOR OF KAPPA LIGHT CHAIN ENHANCER B CELLS INHIBITOR PD028227: M1-L62	BLAST_PRODOM
33	7513245CD1	442	signal_cleavage: M1-A34	SPSCAN
			Signal Peptide: M1-A31, M1-A32, M1-A34, P8-A30	HMMER
			Calcium-binding EGF-like domain: D77-T123, G207-G246, D247-K286	HMMER_SMART

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Low-density lipoprotein receptor domain class A: E46-K83, T85-K124, V126-T165, T169-Q206	HMMER_SMART
			Low-density lipoprotein-receptor YWTD domain: Y313-D355, E360-R402	HMMER_SMART
			Low-density lipoprotein receptor domain: K45-K83, K84-K124, Q125-T165, G168-Q206	HMMER_PFAM
			Low-density lipoprotein receptor repeat: N333-V378, K380-L421	HMMER_PFAM
			Cytosolic domain: M1-P6 Transmembrane domain: G7-L29 Non-cytosolic domain: A30-H442	TMHMMER
			Calcium-binding EGF-like domain IPB001881: Q193-G203, C261-C272	BLIMPS_BLOCKS
			Low density lipoprotein (LDL)-receptor class A (LDLRA) domain IPB002172: C98-E117	BLIMPS_BLOCKS
			Low density lipoprotein (LDL) receptor signature PR00261: G96-E117	BLIMPS_PRINTS
			GLYCOPROTEIN EGF-LIKE DOMAIN APOLIPOPROTEIN E RECEPTOR LIPOPROTEIN APOER-2 DELTA-47 PRECURSOR SIGNAL PD151872: C251-A288	BLAST_PRODOM
			LDL RECEPTOR HOMOLOG LR-8B PROTEIN GLYCOPROTEIN EGF-LIKE DOMAIN PD166320: N333-K380	BLAST_PRODOM
			GLYCOPROTEIN PROTEIN RECEPTOR EGF-LIKE DOMAIN LIPOPROTEIN PRECURSOR SIGNAL TRANSMEMBRANE RECEPTOR RELATED PD149641: R334-G423	BLAST_PRODOM
			HEAD ACTIVATOR BINDING PROTEIN PRECURSOR SIGNAL GLYCOPROTEIN PD175838: P128-C261	BLAST_PRODOM
			EGF-LIKE DOMAIN DM00864|I48952|379-445: E230-F297 DM00864|I48623|338-401: E230-A287	BLAST_DOMO
			LDL RECEPTOR LIGAND-BINDING REPEAT DM00045|P01130|37-111: R58-C127, G96-D173	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			EGF DM00003|P01130|303-357: D197-C251	BLAST_DOMO
			Potential Phosphorylation Sites: S76, S138, S145, S341, T85, T169, T244, T332, T404	MOTIFS
			Potential Glycosylation Sites: N312, N389, N409	MOTIFS
			Cell attachment sequence: R171-D173	MOTIFS
			EGF-like domain signature 2: C231-C245	MOTIFS
			Calcium-binding EGF-like domain pattern signature: D247-C270	MOTIFS
			LDL-receptor class A (LDLRA) domain signature: C59-C81, C98-C122, C141-C163, C182-C204	MOTIFS
34	7513711CD1	82	signal_cleavage: M1-G22	SPSCAN
			Signal Peptide: M1-C16, M1-G19, M1-G21, M1-G22, M1-G23, M1-G24, M1-A26	HMMER
			Cytosolic domain: M1-R6 Transmembrane domain: L7-T29 Non-cytosolic domain: E30-L82	TMHMMER
			INTERLEUKIN-13 RECEPTOR ALPHA-1 CHAIN PRECURSOR IL13R-ALPHA1 IL-13RA1 TRANSMEMBRANE GLYCOPROTEIN PD155860: T29-I78	BLAST_PRODOM
			Potential Glycosylation Sites: N37, N61	MOTIFS
35	7513965CD1	152	signal_cleavage: M1-A21	SPSCAN
			Signal Peptide: M1-A21, M1-L23, M1-L24, M7-A21	HMMER
			Sushi domain (SCR repeat): C24-C82	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Non-cytosolic domain: M1-Q118 Transmembrane domain: A119-W141 Cytosolic domain: Q142-I152	TMHMMER
			EGF-like domain IPB000561: C49-R57	BLIMPS_BLOCKS
			INTERLEUKIN-2 RECEPTOR ALPHA SUBUNIT CHAIN PRECURSOR IL-2 P55 TAC ANTIGEN PD008579: M1-P122	BLAST_PRODOM
			INTERLEUKIN-2 RECEPTOR DM03317|P01589|1-271: M1-P122, C51-I152 DM03317|P41690|1-274: M1-P122, L121-I152, M15-S84 DM03317|P12342|1-274: M1-P122, L121-I152 DM03317|P01590|1-267: M1-L121	BLAST_DOMO
			Potential Phosphorylation Sites: S74, S148, T35, T90, T107	MOTIFS
			Potential Glycosylation Sites: N70, N89	MOTIFS
36	7513969CD1	451	hormone ligand-binding domain in nuclear receptors: K220-E378	HMMER_SMART
			C4 zinc finger in nuclear hormone receptors: D50-D123	HMMER_SMART
			Ligand-binding domain of nuclear hormone receptor: T223-L408	HMMER_PFAM
			Zinc finger, C4 type (two domains): E51-D128	HMMER_PFAM
			C4-type steroid receptor zinc finger IPB001628: C53-I81, Y89-V125	BLIMPS_BLOCKS
			Nuclear hormones receptors DNA-binding region signature: S34-I102	PROFILESCAN
			C4-type steroid receptor zinc finger signature PR00047: C53-T69, T69-N84, R104-F112, F112-M120	BLIMPS_PRINTS
			Thyroid hormone receptor signature PR00546: G41-V55, R79-S90, V118-K134, R135-S153, P158-H175, Q181-P200, F218-F239, C244-A263, F309-G331	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			RECEPTOR THYROID HORMONE PROTEIN TRANSCRIPTION REGULATION DNA-BINDING NUCLEAR ZINC-FINGER MULTIGENE PD002338: M122-A225	BLAST_PRODOM
			THYROID HORMONE RECEPTOR ALPHA-2 TRANSCRIPTION REGULATION DNA-BINDING NUCLEAR PROTEIN ZINC-FINGER PD032276: L368-G421	BLAST_PRODOM
			PROTEIN RECEPTOR NUCLEAR TRANSCRIPTION REGULATION DNA-BINDING ZINC-FINGER HORMONE FAMILY MULTIGENE PD000035: C53-M120	BLAST_PRODOM
			RECEPTOR PROTEIN NUCLEAR TRANSCRIPTION REGULATION DNA-BINDING ZINC-FINGER HORMONE FAMILY MULTIGENE PD000112: F218-Q380	BLAST_PRODOM
			NUCLEAR HORMONES RECEPTORS DNA-BINDING REGION DM00047|S09178|43-335: I43-D336 DM00047|P03373|27-319: I43-D336 DM00047|P37243|112-404: I43-D336 DM00047|P18119|50-342: I43-D336	BLAST_DOMO
			Potential Phosphorylation Sites: S6, S12, S18, S34, S45, S90, S129, S203, S260, S396, S426, S432, S433, S434, T161, T327	MOTIFS
			Nuclear hormones receptors DNA-binding region signature: C53-R79	MOTIFS
37	7512119CD1	399	Non-cytosolic domains: M1-Q105, H158-C171, H254-Q267, R331-G368 Transmembrane domains: W106-M128, N135-T157, V172-E194, F231-M253, T268-V285, T313-M330, V369-L391 Cytosolic domains: K129-M134, F195-F230, K286-D312, N392-K399	TMHMMER
			Potential Phosphorylation Sites: S36, S89, S356, T44, T60, T91, T227, T262, T313, Y66	MOTIFS
			Potential Glycosylation Sites: N34, N135, N203, N373	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
38	7515577CD1	231	signal_cleavage: M1-G40	SPSCAN
			Signal Peptide: M21-A39, M21-G40, M21-T41, M21-P42, M21-V46	HMMER
			Cytosolic domains: M1-T25, R99-R231 Transmembrane domains: L26-I48, D76-L98 Non-cytosolic domain: R49-A75	TMHMMER
			Potential Phosphorylation Sites: S144, T59	MOTIFS
			Potential Glycosylation Sites: N142	MOTIFS
39	7514748CD1	203	signal_cleavage: M1-A21	SPSCAN
			Signal Peptide: M1-A21, Q4-A21, M1-S22, M1-E24, M1-E27, M1-S28	HMMER
			Immunoglobulin: P29-L141	HMMER_SMART
			Immunoglobulin domain: G37-V122	HMMER_PFAM
			V type Ig domains from SCOP (antibody variable domain): L23-G151	HMMER_INCY
			Ig superfamily from SCOP: V25-G129	HMMER_INCY
			CHANNEL SUBUNIT SODIUM IONIC VOLTAGE-GATED IMMUNOGLOBULIN TRANSMEMBRANE DOMAIN ION GLYCOPROTEIN PD013099: I32-G151	BLAST_PRODOM
			Potential Phosphorylation Sites: S3, T139	MOTIFS
			Potential Glycosylation Sites: N102, N108	MOTIFS
40	7513838CD1	525	signal_cleavage: M1-A16	SPSCAN
			Signal Peptide: M1-A16, M1-E19, M1-L22	HMMER
			Fibronectin type 3 domain: P355-L454	HMMER_SMART
			Ephrin receptor ligand binding domain: E18-C217	HMMER_PFAM
			Fibronectin type III domain: P355-S455	HMMER_PFAM
			Laminin-type EGF-like (LE) domain IPB002049: G406-V416	BLIMPS_BLOCKS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Receptor tyrosine kinase class V IPB001426: E18-E43, R87-E121, Q178-Q231, P267-Q291, C299-P346, V365-V391, C409-S452	BLIMPS_BLOCKS
			EPH FAMILY PROTEIN PD002683: P355-E471	BLAST_PRODOM
			KINASE RECEPTOR PRECURSOR TYROSINE PROTEIN EPHRIN TRANSFERASE ATP-BINDING PHOSPHORYLATION TRANSMEMBRANE GLYCOPROTEIN PD001495: E18-F149, G172-C217 PD149648: P218-A301	BLAST_PRODOM
			KINASE RECEPTOR PRECURSOR TYROSINE PROTEIN EPHRIN TRANSFERASE ATP-BINDING PHOSPHORYLATION TRANSMEMBRANE SIGNAL PD001551: C302-P354	BLAST_PRODOM
			RECEPTOR TYROSINE KINASE CLASS V DM00501|P54762|21-373: L21-F149, L177-S403 DM00501|P29317|30-379: V20-S152, L177-C409 DM00501|P29319|31-375: L21-F149, G172-S403, V391-C412 DM00501|P29320|31-376: L21-F149, G172-C395	BLAST_DOMO
			Potential Phosphorylation Sites: S45, S95, S226, T24, T54, T69, T108, T138, T190, T468	MOTIFS
			Potential Glycosylation Sites: N465	MOTIFS
			Cell attachment sequence: R294-D296, R383-D385	MOTIFS
			EGF-like domain signature 2: C284-C299	MOTIFS
41	7515163CD1	217	signal_cleavage: M1-P60	SPSCAN
			Signal Peptide: M1-A31, M3-V27, M3-G29	HMMER
			Immunoglobulin: G35-A144	HMMER_SMART
			Immunoglobulin V-Type: T45-I128	HMMER_SMART
			V type Ig domains from SCOP (antibody variable domain): L23-S150	HMMER_INCY

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Cytosolic domains: M1-S6, K178-N217 Transmembrane domains: A7-G29, A155-W177 Non-cytosolic domain: H30-S154	TMHMMER
			Receptor tyrosine kinase class III IPB001824: E106-A144	BLIMPS_BLOCKS
			BUTYROPHILIN PRECURSOR SIGNAL IMMUNOGLOBULIN PROTEIN TRANSMEMBRANE GLYCOPROTEIN FOLD MYELINOLIGODENDROCYTE MYELIN PD000570: V39-C126	BLAST_PRODOM
			Potential Phosphorylation Sites: S123, S134, S207, Y124, Y188	MOTIFS
			Potential Glycosylation Sites: N132	MOTIFS
42	7516929CD1	790	Non-cytosolic domain: M1-P762 Transmembrane domain: W763-L785 Cytosolic domain: A786-P790	TMHMMER
			Forkhead associated domain: P27-L85	HMMER_SMART
			FHA domain: I28-G106	HMMER_PFAM
			SARCOLEMMAL-ASSOCIATED PROTEIN 2 PROTEIN 3 PD075105: L342-S442	BLAST_PRODOM
			PROTEIN COILED-COIL CHAIN MYOSIN REPEAT HEAVY ATP-BINDING FILAMENT HEPTAD PD000002: L214-Q432, K229-E467, L243-K481, E274-R516, M292-E531, K315-E545, I331-E538, L443-Q662, K458-L701, K479-Q720, L510-R755,	BLAST_PRODOM
			SARCOLEMMAL-ASSOCIATED PROTEIN PROTEIN 2 PROTEIN 3 PD020745: N759-P790	BLAST_PRODOM
			TRICHOHYALIN DM03839|P22793|921-1475: S162-R693, K229-E756, K271-Q746 DM03839|P37709|632-1103: Q167-E627, D209-E655, K255-E717, E283-K757, Q328-K761	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			SIMILAR TO MYOSIN-LIKE PROTEIN MLP-1 DM07884|Q02455|35-1728: V153-K761, Y161-L751, Q167-E756, T157-K761, Q231-E731, S159-E736, E174-N759, E179-K737, R216-K757, E264-K757 DM07884|P40457|16-1657: E183-K757, Q167-L751, E264-E756, V141-Q605, N230-K757, E164-K761	BLAST_DOMO
			Potential Phosphorylation Sites: S82, S148, S215, S235, S257, S273, S280, S318, S420, S453, S503, S534, S541, S558, S566, S586, S637, S659, S676, S686, S703, S739, T9, T64, T113, T117, T124, T253, T282, T296, T477, T535, T733	MOTIFS
			Potential Glycosylation Sites: N43, N76, N230, N400	MOTIFS
43	7515570CD1	230	signal_cleavage: M1-A21	SPSCAN
			Signal Peptide: M1-V17, M1-A21, M1-T23, M1-G24, M1-A29	HMMER
			Cytosolic domains: M1-A6, C107-Y230 Transmembrane domains: L7-A29, I84-I106 Non-cytosolic domain: S30-Y83	TMHMMER
			Legume lectin alpha domain IPB000985: T5-G24	BLIMPS_BLOCKS
			Legume lectins beta IPB001220: T5-G24	BLIMPS_BLOCKS
			Potential Phosphorylation Sites: S28, S30, S118, S147, S162, T23, T62, T70, T119, T154, T171	MOTIFS
			Potential Glycosylation Sites: N160	MOTIFS
44	7515680CD1	196	signal_cleavage: M1-S18	SPSCAN
			Signal Peptide: M1-S18	HMMER
			Signal Peptide: M1-S20	HMMER
			Signal Peptide: M1-G22	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Cytosolic domain: R149-N196 Transmembrane domain: I126-Y148 Non-cytosolic domain: M1-S125	TMHMMER
			Potential Phosphorylation Sites: S56 S66 S168 S172 T26 T95 T119 T177	MOTIFS
			Potential Glycosylation Sites: N19 N33 N39 N99 N113	MOTIFS
45	7516698CD1	367	Signal Peptide: M27-L47	HMMER
			Signal Peptide: M27-S48	HMMER
			Signal Peptide: M27-A50	HMMER
			Cytosolic domains: M1-Q31, K76-R95, P237-N292 Transmembrane domains: H32-F54, I58-S75, N96-F115, L214-G236, L293-Y315 Non-cytosolic domains: L55-S57, H116-K213, L316-S367	TMHMMER
			Potential Phosphorylation Sites: S6 S13 S20 S159 S189 T202 T286	MOTIFS
			Potential Glycosylation Sites: N168 N187	MOTIFS
			Trp-Asp (WD) repeats signature: I3-I17	MOTIFS
46	7517501CD1	815	signal_cleavage: M1-G17	SPSCAN
			Signal Peptide: M1-G19	HMMER
			Signal Peptide: M1-P21	HMMER
			Signal Peptide: M1-V22	HMMER
			Immunoglobulin: V27-K102, P209-V296, A112-Q196, N308-R399, V404-G504	HMMER_SMART
			Immunoglobulin C-2 Type: I215-G285, F410-G492	HMMER_SMART
			Tyrosine kinase, catalytic domain: L582-V807	HMMER_SMART
			Immunoglobulin domain: G217-A280, N412-A487, G35-E86, G316-A383	HMMER_PFAM
			Ig superfamily from SCOP: I23-P107, L205-L301, S304-R390, Y400-G494, W108-I199	HMMER_INCY
			Cytosolic domain: Y538-A815 Transmembrane domain: F515-L537 Non-cytosolic domain: M1-L514	TMHMMER
			Receptor tyrosine kinase class III IPB001824: P21-P53, A95-P104, V323-Q332, R361-R399, P570-A622, H623-R677, D754-V793	BLIMPS_BLOCKS
			Receptor tyrosine kinase class III signature: E395-Y454	PROFILESCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Tyrosine kinase catalytic domain signature PR00109: T663-R676	BLIMPS_PRINTS
			RECEPTOR KINASE PRECURSOR TYROSINE PROTEIN TRANSFERASE TRANSMEMBRANE GLYCOPROTEIN PHOSPHORYLATION FACTOR GROWTH PD003025: A391-L582	BLAST_PRODOM
			RECEPTOR FMS TYROSINE PROTEIN KINASE TRANSMEMBRANE TRANSFERASE GLYCOPROTEIN PHOSPHORYLATION MACROPHAGE COLONY PD069181: M1-C84	BLAST_PRODOM
			RECEPTOR FMS TYROSINE PROTEIN KINASE TRANSMEMBRANE TRANSFERASE GLYCOPROTEIN PHOSPHORYLATION MACROPHAGE COLONY PD012171: K678-F737	BLAST_PRODOM
			RECEPTOR FACTOR KINASE GROWTH PRECURSOR TYROSINE PROTEIN TRANSMEMBRANE TRANSFERASE PHOSPHORYLATION GLYCOPROTEIN PD003030: T85-V279	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM00004|P13369|580-900: Q583-N773	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004|Q00495|581-900: Q583-N773	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004|P09581|581-899: Q583-N773	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004|P04048|28-354: F584-N773	BLAST_DOMO
			Potential Phosphorylation Sites: S26 S190 S227 S713 S732 S736 S738 T39 T81 T314 T355 T359 T397 T438 T562 T722 Y300 Y379	MOTIFS
			Potential Glycosylation Sites: N45 N73 N153 N240 N275 N302 N335 N353 N412 N428 N480 N734	MOTIFS
			Protein kinases ATP-binding region signature: L588-K612	MOTIFS
			Receptor tyrosine kinase class III signature: G641-T654	MOTIFS
47	7518576CD1	260	signal_cleavage: M1-A30	SPSCAN
			Signal Peptide: L11-A30	HMMER
			Signal Peptide: M5-A30	HMMER
			Signal Peptide: L14-G32	HMMER
			Signal Peptide: M5-G32	HMMER
			Immunoglobulin: Q83-K182	HMMER_SMART
			Immunoglobulin domain: G91-K162	HMMER_PFAM
			V type Ig domains from SCOP: W78-S188	HMMER_INCY

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Ig superfamily from SCOP: I79-S185	HMMER_INCY
			Cytosolic domain: A230-E260 Transmembrane domain: L207-L229 Non-cytosolic domain: M1-L206	TMHMMER
			Potential Phosphorylation Sites: S34 S35 S65 S139 S151 S178 S187 S232 T89 T231	MOTIFS
			Potential Glycosylation Sites: N97 N149	MOTIFS
48	7518626CD1	237	signal_cleavage: M1-G30	SPSCAN
			Signal Peptide: M6-G28	HMMER
			Signal Peptide: M6-G30	HMMER
			Signal Peptide: M6-A35	HMMER
			Low-density lipoprotein receptor domain cla: S53-E91, R93-T124	HMMER_SMART
			Low-density lipoprotein receptor domain: G52-E91, L92-T124	HMMER_PFAM
			Cytosolic domains: M1-Q8, R209-P237 Transmembrane domains: V9-L31, V186-L208 Non-cytosolic domain: E32-G185	TMHMMER
			Low density lipoprotein (LDL)-receptor class A (LDLRA) domain IPB002172: C100-E119	BLIMPS_BLOCKS
			4-disulphide core signature PR00003: A47-P56, R73-C80, G121-P129	BLIMPS_PRINTS
			Low density lipoprotein (LDL) receptor signature PR00261: G65-E86	BLIMPS_PRINTS
			GLYCOPROTEIN RECEPTOR EGF-LIKE DOMAIN PRECURSOR LIPOPROTEIN SIGNAL TRANSMEMBRANE PROTEIN REPEAT PD000144: S51-T95	BLAST_PRODOM
			LDL RECEPTOR LIGAND-BINDING REPEAT DM00045|P01130|37-111: G65-C122	BLAST_DOMO
			Potential Phosphorylation Sites: S84 S116 S147 S173 T71 T95 T104	MOTIFS
			Potential Glycosylation Sites: N150 N168	MOTIFS
			LDL-receptor class A (LDLRA) domain signature: C67-C89, C100-C122	MOTIFS
49	7515714CD1	170	signal_cleavage: M1-G28	SPSCAN
			Signal Peptide: P4-G28	HMMER
			Signal Peptide: M1-G28	HMMER
			Signal Peptide: A5-G28	HMMER
			Immunoglobulin: E38-K133	HMMER_SMART

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Immunoglobulin domain: G46-K121	HMMER_PFAM
			SIGNAL REGULATORY PROTEIN BETA1 PRECURSOR SIRP-BETA1 SIGNAL IMMUNOGLOBULIN FOLD GLYCOPROTEIN TRANSMEMBRANE PD087237: M1-V43	BLAST_PRODOM
			Potential Phosphorylation Sites: S103 S126 S161 S167 T90 T95 T110	MOTIFS

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
50/2847449CB1 968	1-218, 1-260, 1-284, 1-553, 1-651, 15-404, 301-807, 325-604, 364-567, 373-968, 383-668
51/7523642CB1 1891	1-756, 1-860, 2-1890, 226-1233, 506-1232, 677-1424, 682-1412, 1031-1891, 1096-1891, 1119-1891, 1246-1891, 1255-1891, 1290-1891
52/7521994CB1 1171	1-686, 2-1170, 212-1171, 227-1171, 306-1171, 345-1171, 545-1171, 559-1171, 571-1171, 975-1171
53/7522289CB1 638	1-638, 2-638, 3-637, 6-638
54/7522336CB1 571	1-248, 1-571
55/7522339CB1 433	1-429, 1-433, 2-432
56/7522361CB1 619	1-262, 1-557, 1-619, 2-618, 284-409
57/7522368CB1 561	1-561, 2-560, 122-561
58/7522373CB1 724	1-632, 2-721, 7-724, 10-724, 65-724, 80-724, 108-724, 115-724, 125-724, 131-724, 137-186, 147-184, 206-724, 245-721, 279-724, 308-724, 310-724, 367-724, 449-724
59/7522381CB1 703	1-683, 2-610, 97-703
60/7523596CB1 1722	1-689, 1-721, 1-884, 2-1721, 24-986, 38-971, 579-1517, 602-1513, 828-1722, 901-1722, 1077-1722, 1086-1722, 1121-1722
61/7523643CB1 1685	1-795, 2-1684, 170-1063, 294-1065, 553-1312, 578-1414, 698-788, 776-866, 776-869, 992-1685
62/7523769CB1 1578	1-733, 1-846, 2-1577, 19-938, 661-1578, 675-1578, 699-1578, 726-1578, 834-1578, 893-1578, 933-1578, 942-1578, 977-1578
63/7523785CB1 1349	1-642, 1-763, 2-765, 2-1349, 27-695, 124-793, 489-1349, 554-1349, 577-1349, 580-1349, 606-1285, 683-1349, 704-1349, 713-1349, 748-1349

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
64/7523836CB1 1709	1-813, 2-1708, 130-1035, 139-1010, 213-1034, 307-1000, 583-1415, 605-1452, 605-1566, 620-1323, 925-1709
65/7523879CB1 1760	1-563, 1-800, 2-1759, 143-1160, 210-1135, 733-1546, 735-1561, 833-1760, 957-1760, 998-1759, 1001-1759
66/7523880CB1 1636	1-741, 2-1634, 285-1111, 389-1111, 584-1296, 587-1465, 754-1636, 864-1635, 886-1636
67/7523812CB1 2850	1-642, 2-2849, 157-2849, 564-1432, 584-1144, 756-1471, 758-1466, 1337-1978, 1384-2156, 1507-2180, 2052-2850
68/7524026CB1 1136	1-659, 1-707, 1-750, 1-753, 1-755, 2-1135, 404-1136
69/7524357CB1 1508	1-626, 1-797, 1-824, 2-1507, 344-825, 534-996, 536-1280, 549-1381, 549-1442, 708-1508
70/7524808CB1 1258	1-688, 2-687, 2-1258, 3-687, 380-1258, 394-430, 433-469
71/7522161CB1 1396	1-1396, 129-774, 129-910, 142-869, 142-881, 659-1285, 990-1396
72/7523999CB1 4466	1-588, 1-655, 1-716, 1-763, 2-4417, 511-1287, 545-1293, 899-1764, 902-1762, 1641-2446, 1759-2444, 2343-3185, 2600-3184, 3008-3864, 3052-3869, 3615-4466, 3735-4466, 3826-4466
73/7524024CB1 2550	1-720, 1-760, 1-809, 1-819, 1-2550, 764-1678, 768-1527, 782-1609, 1183-1901, 1207-1904, 1784-2550, 1841-2550, 1858-2550
74/7522455CB1 1352	1-694, 1-748, 1-764, 1-782, 1-791, 1-796, 1-804, 1-1351, 2-1351, 11-757, 534-1352, 553-1351, 643-1352, 724-1352
75/7524494CB1 4242	1-893, 6-903, 12-970, 60-765, 60-905, 106-4242, 710-1483, 718-1534, 718-1565, 718-1574, 719-1490, 734-1483, 736-1438, 736-1444, 736-1567, 741-1484, 762-1384, 1479-2377, 1480-2403, 1905-2804, 2037-2817, 2745-3588, 2747-3587, 2755-3551, 2756-3587, 2785-3551, 2794-3548, 2794-3588, 2830-3551, 2861-3554, 2864-3554, 3066-3548, 3394-4242, 3481-4242, 3482-4242, 3523-4242, 3529-4242
76/7524965CB1 1584	1-718, 1-775, 1-880, 2-1529, 660-1584, 735-1584, 928-1584

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
77/7525018CB1 2239	1-593, 1-631, 1-670, 1-687, 1-698, 6-2239, 594-1372, 594-1490, 690-1612, 806-1606, 1493-2239
78/7516620CB1 6929	1-6929, 693-1218, 709-1289, 866-1146, 1001-1539, 1150-1407, 1258-1887, 1960-2223, 2291-2502, 2410-3044, 2634-3178, 2748-3469, 2772-3150, 2794-3265, 3118-3397, 3160-4030, 3206-4032, 3224-4033, 3331-3746, 3345-3536, 3348-3928, 3414-3669, 3416-4033, 3466-3651, 3517-3991, 3530-4146, 3547-4334, 3776-4070, 3781-4308, 3830-4384, 3850-4146, 3866-4100, 3894-4643, 3928-4457, 3938-4188, 3942-4554, 3960-4257, 4008-4491, 4012-4263, 4041-4584, 4255-4905, 4343-4928, 4384-4621, 4406-5015, 4470-5066, 4476-4772, 4519-5069, 4540-5123, 4553-5159, 4658-4791, 4700-4954, 4703-5269, 4753-5007, 4753-5209, 4776-4874, 4776-5028, 4822-5030, 4836-5244, 4836-5334, 4849-4954, 4851-4954, 4888-5226, 4902-5125, 4902-5135, 4902-5138, 4925-5208, 4928-5202, 4936-5197, 4964-5189, 5031-5315, 5301-5769, 5538-5966, 5562-5998, 5562-6059, 5564-6094, 5574-5834, 5581-5810, 5581-5813, 5581-6057, 5587-5831, 5587-5856, 5604-5863, 5610-5839, 5625-6174, 5628-5876, 5638-6172, 5661-5802, 5663-5910, 5669-6229, 5680-5931, 5688-5952, 5691-6085, 5704-6229,
	5712-6302, 5722-5929, 5724-6328, 5731-5973, 5731-5984, 5755-6179, 5765-6336, 5767-6062, 5830-6331, 5852-6229, 5859-6090, 5890-6064, 5898-6289, 5936-6114, 5944-6181, 5978-6329, 5981-6260, 5994-6095, 6023-6317, 6026-6204, 6045-6342, 6054-6601, 6055-6332, 6080-6319, 6119-6326, 6143-6352, 6144-6339, 6164-6683, 6174-6443, 6204-6432, 6211-6518, 6215-6471, 6257-6534, 6275-6541, 6281-6549, 6293-6516, 6304-6926, 6310-6869, 6311-6906, 6317-6928, 6320-6577, 6320-6608, 6320-6613, 6323-6626, 6329-6568, 6358-6611, 6364-6794, 6377-6610, 6377-6905, 6382-6721, 6388-6929, 6427-6680, 6427-6911, 6429-6925, 6436-6642, 6443-6656, 6443-6704, 6462-6929, 6466-6926, 6472-6929, 6482-6929, 6493-6874, 6497-6926, 6504-6927, 6504-6929, 6507-6755, 6508-6771, 6509-6926, 6511-6652, 6516-6926, 6517-6926, 6519-6925, 6521-6867, 6521-6929, 6533-6929, 6541-6813, 6542-6929, 6544-6765, 6545-6929, 6550-6866, 6556-6925, 6561-6925, 6566-6856, 6568-6929, 6569-6913, 6570-6849, 6571-6851, 6589-6830, 6601-6926, 6615-6929, 6631-6929, 6649-6926, 6668-6929,
	6677-6929, 6705-6929, 6711-6925, 6724-6928, 6737-6926, 6741-6929, 6748-6929, 6753-6926, 6778-6926, 6779-6918, 6779-6919, 6831-6929, 6834-6929
79/7525149CB1 2273	1-629, 1-682, 1-899, 2-2272, 487-1285, 492-1280, 838-1755, 1025-1755, 1427-2268, 1530-2273, 1646-2273, 1988-2043, 1988-2049, 2080-2141

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
80/7513047CB1 697	1-201, 1-242, 12-241, 12-246, 14-242, 21-697, 58-324, 60-239, 65-241, 65-243, 120-194, 238-697, 243-472, 243-494, 243-503, 243-507, 243-554, 243-559, 243-697, 244-482, 245-697, 251-529, 251-536, 251-675, 254-514, 272-697, 283-697, 285-566, 286-687, 289-697, 297-675, 298-552, 300-686, 316-697, 322-604, 322-612, 334-576, 334-626, 335-660, 341-530, 342-689, 345-583, 352-687, 353-689, 366-687, 368-688, 387-686, 398-689, 400-623, 413-549, 415-697, 448-689, 462-676, 462-696, 464-697, 472-675, 476-697, 479-697, 482-697, 489-697, 505-691, 512-697, 544-694, 553-691, 565-685, 565-692, 569-687, 594-689, 604-690, 605-687, 615-687, 629-697
81/7513056CB1 1150	1-239, 11-239, 12-239, 12-1128, 19-734, 34-239, 55-289, 55-331, 59-281, 62-190, 189-577, 239-463, 246-525, 283-893, 284-847, 285-398, 292-425, 305-913, 336-584, 339-932, 359-896, 361-520, 380-1032, 380-1037, 387-681, 389-616, 389-874, 407-699, 435-952, 445-518, 494-1060, 501-1079, 523-725, 523-737, 523-1043, 523-1125, 524-1026, 528-756, 532-831, 560-991, 561-866, 571-911, 571-1086, 572-1118, 574-821, 584-875, 613-952, 620-1150, 629-1141, 632-876, 647-1015, 653-1072, 658-1130, 664-1127, 671-928, 673-1128, 674-1128, 676-1128, 677-1129, 681-1143, 688-1113, 700-1129, 700-1140, 703-1128, 721-1136, 733-1131, 734-1127, 736-1127, 741-1128, 755-1036, 765-930, 770-1124, 774-1081, 784-1134, 793-1129, 805-1140, 818-1134, 858-1104, 872-998, 910-1133, 914-1129, 915-1148, 930-1148, 938-1073, 974-1142, 1002-1148, 1007-1128, 1011-1134, 1011-1150
82/7513245CB1 2459	1-451, 18-2459, 95-504, 102-1028, 102-1052, 876-1474, 979-1259, 998-1198, 1284-2189, 1291-1462, 1347-1761, 1439-1704, 1453-1720, 1491-1580, 1525-1881, 2084-2459, 2087-2340, 2114-2337, 2123-2427, 2128-2453, 2136-2451, 2171-2392, 2172-2451, 2245-2459, 2254-2451, 2275-2430, 2338-2454
83/7513711CB1 1389	1-945, 701-1389
84/7513965CB1 1220	1-1220, 197-708
85/7513969CB1 1854	1-550, 1-551, 1-572, 1-618, 1-619, 1-645, 1-673, 1-678, 1-693, 1-781, 1-788, 1-812, 1-841, 2-1854, 665-1404, 724-868, 940-1404
86/7512119CB1 1885	1-899, 431-1335, 617-1594, 1150-1885
87/7515577CB1 801	1-751, 1-801, 2-800, 22-801
88/7514748CB1 1344	1-1344, 238-1092, 468-1297

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
89/7513838CB1 3714	1-249, 1-527, 1-551, 1-3714, 3-550, 5-330, 8-672, 46-651, 64-886, 65-667, 65-676, 65-892, 65-911, 65-1003, 66-330, 66-624, 66-659, 68-658, 72-906, 108-550, 122-695, 138-594, 197-481, 197-714, 252-446, 252-451, 252-923, 305-534, 305-661, 501-796, 541-1215, 556-661, 701-1235, 708-1168, 755-996, 774-1515, 912-1570, 948-1328, 965-1332, 965-1543, 969-1302, 969-1427, 1049-1683, 1129-1369, 1143-1326, 1143-1376, 1143-1660, 1149-1443, 1187-1428, 1232-1853, 1277-1642, 1298-1854, 1343-1569, 1484-1751, 1536-1837, 1536-1930, 1560-1878, 1603-1880, 1644-1902, 1693-1945, 3523-3565, 3542-3578, 3544-3577
90/7515163CB1 1447	1-812, 1-815, 1-816, 2-771, 685-1447
91/7516929CB1 3136	1-622, 118-658, 118-1648, 125-611, 309-789, 424-1037, 558-1053, 633-1268, 700-1169, 701-849, 764-987, 764-1016, 764-1067, 764-1080, 770-1405, 777-1084, 922-1072, 948-1039, 948-1224, 1078-1552, 1108-1552, 1123-1490, 1125-1552, 1159-1569, 1164-1611, 1191-1552, 1273-1535, 1279-1415, 1394-1493, 1510-2044, 1647-1897, 1647-1930, 1647-1947, 1647-2043, 1647-2121, 1647-2204, 1647-2224, 1647-2236, 1647-2265, 1647-2280, 1648-2355, 1652-2285, 1662-1920, 1662-1943, 1662-2566, 1666-1947, 1671-2220, 1672-2235, 1675-1934, 1682-2198, 1684-2399, 1684-2435, 1684-2457, 1684-2474, 1684-2547, 1684-2554, 1684-2650, 1685-2572, 1709-1933, 1709-1938, 1709-2054, 1709-2194, 1716-2427, 1723-2351, 1724-2663, 1741-2658, 1767-2300, 1813-2331, 1816-2687, 1817-2563, 1827-2702, 1830-2392, 1834-2657, 1857-2691, 1893-2863, 1916-2696, 1917-2173, 1920-2804, 1956-2798, 1977-2697, 1987-2881, 2000-2555, 2000-2625, 2000-2628, 2000-2650, 2003-2442, 2003-2507, 2003-2509, 2003-2552, 2003-2590, 2003-2650, 2003-2657, 2003-2673, 2003-2681, 2003-2687,
	2003-2695, 2003-2710, 2003-2711, 2003-2745, 2003-2781, 2003-2787, 2003-2837, 2003-2838, 2003-2839, 2003-2847, 2003-2858, 2003-2886, 2003-2911, 2003-2917, 2003-2940, 2003-2965, 2003-2975, 2003-2987, 2005-2943, 2019-2971, 2048-2557, 2059-2486, 2091-3136, 2092-2887, 2102-2945, 2141-2745, 2163-2669, 2224-2597, 2226-3083, 2245-2599, 2314-2739, 2314-2742, 2350-2617, 2350-2635, 2361-2907, 2430-2964
92/7515570CB1 840	1-100, 1-219, 1-702, 1-784, 1-838, 1-840, 2-839, 2-840, 4-840
93/7515680CB1 621	1-621, 3-620
94/7516698CB1 1349	1-806, 579-1349

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
95/7517501CB1 3868	1-106, 1-195, 1-204, 1-206, 1-210, 1-232, 1-245, 1-396, 1-429, 1-452, 1-584, 1-586, 1-645, 1-749, 1-3868, 2-380, 4-637, 5-520, 8-329, 8-474, 8-510, 8-556, 8-580, 14-332, 21-269, 31-585, 39-227, 39-554, 60-510, 71-330, 72-506, 79-508, 83-585, 97-533, 107-263, 107-675, 129-681, 137-587, 152-586, 160-554, 160-703, 172-391, 183-762, 189-429, 189-586, 303-682, 331-587, 331-748, 378-627, 390-648, 391-906, 424-1025, 450-587, 534-1109, 555-682, 555-801, 556-1128, 572-1110, 573-1026, 588-910, 604-1115, 624-809, 649-748, 649-910, 649-1057, 664-1127, 676-1138, 706-910, 706-1061, 706-1198, 707-1110, 714-1219, 714-1317, 728-1026, 728-1110, 749-1051, 749-1109, 749-1310, 750-1110, 750-1122, 750-1306, 777-1313, 790-1026, 802-1371, 830-1074, 838-1482, 858-1133, 889-1101, 911-1061, 911-1347, 911-1350, 959-1225, 1005-1251, 1045-1515, 1107-1247, 1107-1376, 1111-1365, 1134-1769, 1161-1429, 1199-1310, 1199-1481, 1199-1585, 1203-1310, 1204-1758, 1220-1310, 1220-1379, 1236-1455, 1236-1503, 1315-1456, 1316-1610, 1380-1585,
	1403-1776, 1426-1659, 1444-1708, 1444-1764, 1456-1585, 1487-1936, 1530-2159, 1554-2224, 1581-1802, 1586-1776, 1598-2159, 1624-1896, 1626-1860, 1626-1861, 1652-1776, 1657-2155, 1692-1992, 1697-1953, 1704-1915, 1706-2155, 1747-2041, 1750-2151, 1752-2066, 1757-2159, 1777-1912, 1794-2044, 1794-2053, 1806-2087, 1841-2026, 1854-2113, 1867-2149, 1871-2125, 1874-2107, 1874-2220, 1874-2351, 1899-2159, 1941-2022, 1958-2141, 1974-2231, 1975-2123, 1975-2300, 1987-2155, 2001-2256, 2043-2220, 2078-2275, 2079-2288, 2092-2296, 2092-2299, 2114-2351, 2129-2391, 2130-2316, 2142-2359, 2164-2391, 2167-2391, 2192-2451, 2240-2378, 2254-2371, 2371-3004, 2503-2750, 2521-3009, 2522-2789, 2547-2803, 2547-3066, 2548-2750, 2567-2728, 2571-2852, 2584-2832, 2587-2868, 2608-2877, 2612-2953, 2615-2872, 2627-2892, 2627-2908, 2634-2843, 2640-3193, 2651-2912, 2657-2913, 2678-2883, 2685-2898, 2688-2935, 2699-2804, 2708-2943, 2713-2985, 2717-2930, 2719-2974, 2719-2994, 2719-3016, 2729-2843, 2782-3056, 2783-2934, 2784-3042, 2784-3069, 2785-3013,
	2785-3050, 2788-2953, 2790-3047, 2790-3217, 2799-3067, 2804-3001, 2838-3075, 2844-3009, 2881-3161, 2896-3446, 2909-3178, 2911-3165, 2913-3067, 2920-3173, 2920-3402, 2923-3140, 2926-3408, 2926-3637, 2931-3195, 2932-3199, 2933-3189, 2935-3219, 2950-3543, 2952-3271, 2962-3181, 2996-3204, 2996-3558, 3004-3222, 3010-3204, 3010-3214, 3028-3309, 3033-3250, 3033-3822, 3051-3320, 3057-3270, 3064-3343, 3068-3209, 3068-3446, 3088-3248, 3088-3326, 3088-3343, 3088-3383, 3100-3381, 3100-3440, 3104-3400, 3110-3679, 3118-3754, 3123-3298, 3128-3348, 3134-3744, 3144-3378, 3148-3427, 3151-3398, 3153-3421, 3162-3410, 3164-3701, 3166-3756, 3167-3730, 3171-3765, 3173-3435, 3173-3672, 3187-3868, 3192-3396, 3193-3867, 3196-3411, 3202-3472, 3204-3454, 3210-3479, 3210-3638, 3215-3350, 3215-3429, 3217-3432, 3217-3480, 3219-3507, 3219-3732, 3234-3726, 3234-3758, 3236-3520, 3245-3868, 3246-3540, 3253-3505, 3258-3577, 3260-3754, 3261-3861, 3273-3868, 3276-3620, 3276-3844, 3277-3757, 3284-3791, 3285-3529, 3285-3820, 3295-3543, 3302-3763,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	3302-3866, 3306-3547, 3306-3637, 3310-3584, 3311-3446, 3313-3786, 3318-3579, 3319-3821, 3333-3562, 3333-3617, 3333-3808, 3333-3816, 3333-3820, 3334-3868, 3335-3673, 3337-3868, 3338-3857, 3338-3865, 3343-3590, 3347-3860, 3347-3861, 3350-3862, 3351-3690, 3352-3601, 3352-3606, 3352-3868, 3357-3578, 3357-3614, 3362-3868, 3364-3829, 3365-3621, 3366-3669, 3368-3624, 3370-3673, 3371-3819, 3374-3645, 3376-3599, 3376-3868, 3379-3577, 3381-3601, 3384-3633, 3384-3684, 3385-3587, 3395-3608, 3395-3639, 3395-3862, 3397-3654, 3397-3657, 3397-3865, 3400-3662, 3401-3685, 3401-3868, 3402-3699, 3402-3865, 3405-3693, 3408-3633, 3408-3864, 3409-3865, 3412-3868, 3417-3668, 3417-3685, 3417-3688, 3417-3868, 3420-3699, 3421-3806, 3422-3868, 3435-3868, 3436-3558, 3436-3667, 3436-3682, 3436-3717, 3436-3719, 3436-3809, 3436-3838, 3436-3851, 3436-3868, 3437-3868, 3443-3697, 3443-3868, 3451-3868, 3452-3784, 3455-3712, 3455-3716, 3455-3868, 3457-3866, 3460-3866, 3464-3705, 3468-3866, 3469-3866, 3470-3703, 3472-3868, 3473-3868, 3474-3801,
	3476-3803, 3479-3868, 3482-3638, 3482-3763, 3483-3868, 3486-3739, 3487-3866, 3489-3868, 3495-3804, 3495-3868, 3501-3753, 3501-3775, 3501-3868, 3504-3734, 3508-3719, 3508-3860, 3520-3780, 3523-3868, 3530-3868, 3544-3867, 3548-3771, 3551-3778, 3555-3788, 3561-3758, 3561-3844, 3566-3848, 3567-3868, 3568-3856, 3591-3800, 3591-3865, 3592-3832, 3595-3840, 3596-3818, 3596-3844, 3596-3866, 3598-3868, 3602-3866, 3608-3860, 3610-3865, 3639-3865, 3645-3868, 3649-3867, 3649-3868, 3650-3867, 3676-3868, 3687-3865, 3690-3868, 3694-3868, 3701-3868, 3724-3822, 3724-3865, 3724-3868, 3746-3868, 3752-3865, 3757-3865, 3791-3865
96/7518576CB1 941	1-184, 1-929, 2-714, 353-693, 380-941
97/7518626CB1 715	1-663, 1-713, 1-715, 2-713, 2-714, 4-713, 13-715, 396-457, 450-511
98/7515714CB1 553	1-552, 1-553

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
50	2847449CB1	DRGCNOT02
78	7516620CB1	FIBRTXS07
80	7513047CB1	BRSTNOT04
81	7513056CB1	UTRSNOT05
82	7513245CB1	HNT2NOT01
89	7513838CB1	THYMNOR02
91	7516929CB1	UTRSTMR01
95	7517501CB1	PLACNOT02

Table 6

Library	Vector	Library Description
BRSTNOT04	PSPORT1	Library was constructed using RNA isolated from breast tissue removed from a 62-year-old East Indian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated an invasive grade 3 ductal carcinoma. Patient history included benign hypertension, hyperlipidemia, and hematuria. Family history included cerebrovascular and cardiovascular disease, hyperlipidemia, and liver cancer.
DRGCNOT02	pINCY	Library was constructed using RNA isolated from dorsal root ganglion tissue removed from the cervical spine of a 32-year-old Caucasian male who died from acute pulmonary edema, acute bronchopneumonia, bilateral pleural effusions, pericardial effusion, and malignant lymphoma (natural killer cell type). The patient presented with pyrexia of unknown origin, malaise, fatigue, and gastrointestinal bleeding. Patient history included probable cytomegalovirus infection, liver congestion, and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, respiratory failure, pneumonia of the left lung, natural killer cell lymphoma of the pharynx, Bell's palsy, and tobacco and alcohol abuse. Previous surgeries included colonoscopy, closed colon biopsy, adenotonsillectomy, and nasopharyngeal endoscopy and biopsy. Patient medications included Diflucan (fluconazole), Deltasone (prednisone), hydrocodone, Lortab, Alprazolam, Reazodone, ProMace-Cytabom, Etoposide, Cisplatin, Cytarabine, and dexamethasome. The patient
		received radiation therapy and multiple blood transfusions.
FIBRTXS07	pINCY	This subtracted library was constructed using 1.3 million clones from a dermal fibroblast library and was subjected to two rounds of subtraction hybridization with 2.8 million clones from an untreated dermal fibroblast tissue library. The starting library for subtraction was constructed using RNA isolated from treated dermal fibroblast tissue removed from the breast of a 31-year-old Caucasian female. The cells were treated with 9CIS retinoic acid. The hybridization probe for subtraction was derived from a similarly constructed library from RNA isolated from untreated dermal fibroblast tissue from the same donor. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR (1991) 19:1954 and Bonaldo, et al., Genome Research (1996) 6:791.
HNT2NOT01	PBLUESCRIPT	Library was constructed at Stratagene (STR937230), using RNA isolated from the hNT2 cell line (derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor).
PLACNOT02	pINCY	Library was constructed using RNA isolated from the placental tissue of a Hispanic female fetus, who was prematurely delivered at 21 weeks' gestation. Serologies of the mother's blood were positive for CMV (cytomegalovirus).

Table 6

Library	Vector	Library Description
THYMNOR02	pINCY	The library was constructed using RNA isolated from thymus tissue removed from a 2-year-old Caucasian female during a thymectomy and patch closure of left atrioventricular fistula. Pathology indicated there was no gross abnormality of the thymus. The patient presented with congenital heart abnormalities. Patient history included double inlet left ventricle and a rudimentary right ventricle, pulmonary hypertension, cyanosis, subaortic stenosis, seizures, and a fracture of the skull base. Family history included reflux neuropathy.
UTRSNOT05	pINCY	The library was constructed using RNA isolated from the uterine tissue of a 45-year-old Caucasian female during a total abdominal hysterectomy and total colectomy. Pathology for the associated tumor tissue indicated multiple leiomyomas of the myometrium and a grade 2 colonic adenocarcinoma of the cecum. Patient history included multiple sclerosis and mitral valve disorder. Family history included type I diabetes, cerebrovascular disease, atherosclerotic coronary artery disease, malignant skin neoplasm, hypertension, and malignant neoplasm of the colon.
UTRSTMR01	pINCY	Library was constructed using RNA isolated from uterine myometrial tissue removed from a 41-year-old Caucasian female during a vaginal hysterectomy. The endometrium was secretory and contained fragments of endometrial polyps. Pathology for associated tumor tissue indicated uterine leiomyoma. Patient history included ventral hernia and a benign ovarian neoplasm.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value = 1.0E-8 or less; Full Length sequences: Probability value = 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value = 1.06E-6; Assembled ESTs: fasta Identity = 95% or greater and Match length = 200 bases or greater; fastx E value = 1.0E-8 or less; Full Length sequences: fastx score = 100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value = 1.0E-3 or less

Table 7

Program	Description	Reference	Parameter Threshold
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM, INCY, SMART or TIGRFAM hits: Probability value = 1.0E-3 or less; Signal peptide hits: Score = 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score ≥ GCG specified "HIGH" value for that particular Prosite motif. Generally, score = 1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score = 120 or greater; Match length = 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score = 3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	

Table 7

Program	Description	Reference	Parameter Threshold
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
50	2847449	2847449F6	SNP00042469	3	17	C	C	T	noncoding	0.53	n/a	n/a	n/a
51	7523642	1733290H1	SNP00019555	210	795	T	T	C	noncoding	0.63	n/a	n/a	n/a
51	7523642	3294566F6	SNP00053413	103	69	C	C	G	A8	n/d	n/a	n/a	n/a
51	7523642	3294839H1	SNP00099838	197	90	G	G	A	R15	0.72	0.81	n/d	0.74
51	7523642	6449492H1	SNP00132265	420	546	A	G	A	noncoding	n/a	n/a	n/a	n/a
51	7523642	6449492H1	SNP00132266	513	639	C	C	T	noncoding	n/a	n/a	n/a	n/a
51	7523642	6576488H1	SNP00019556	427	897	T	T	C	noncoding	0.87	0.79	0.99	0.84
51	7523642	6576488H1	SNP00019557	468	938	G	G	A	noncoding	0.74	0.8	0.9	0.75
51	7523642	6958246H1	SNP00053414	226	651	T	C	T	noncoding	n/a	n/a	n/a	n/a
51	7523642	6958246H1	SNP00073871	85	510	A	A	G	noncoding	0.77	n/a	n/a	n/a
51	7523642	7757655H1	SNP00053414	580	652	C	C	T	noncoding	n/a	n/a	n/a	n/a
55	7522339	2925805T6	SNP00039916	172	288	G	G	A	K93	0.96	0.91	n/d	n/d
56	7522361	1215383H1	SNP00039915	149	161	G	G	C	C51	n/d	n/a	n/a	n/a
56	7522361	2615658F6	SNP00039915	199	152	G	G	C	S48	n/d	n/a	n/a	n/a
56	7522361	3190928H1	SNP00039916	86	474	A	G	A	K155	0.96	0.91	n/d	n/d
60	7523596	1733290H1	SNP00019555	210	535	T	T	C	H163	0.63	n/a	n/a	n/a
60	7523596	3294566F6	SNP00053413	103	69	C	C	G	A8	n/d	n/a	n/a	n/a
60	7523596	3294839H1	SNP00099838	197	90	G	G	A	R15	0.72	0.81	n/d	0.74
60	7523596	6449492H1	SNP00040156	231	97	A	C	A	stop17	0.88	0.92	n/d	0.86
60	7523596	6449492H1	SNP00132265	420	286	A	G	A	A80	n/a	n/a	n/a	n/a
60	7523596	6449492H1	SNP00132266	513	379	C	C	T	L111	n/a	n/a	n/a	n/a
60	7523596	6576488H1	SNP00019556	427	637	T	T	C	D197	0.87	0.79	0.99	0.84
60	7523596	6576488H1	SNP00019557	468	678	G	G	A	R211	0.74	0.8	0.9	0.75
60	7523596	6958246H1	SNP00053414	226	391	T	C	T	N115	n/a	n/a	n/a	n/a
60	7523596	6958246H1	SNP00073871	85	250	A	A	G	V68	0.77	n/a	n/a	n/a
60	7523596	7757655H1	SNP00053414	580	392	C	C	T	L116	n/a	n/a	n/a	n/a
61	7523643	7260543H1	SNP00140016	196	497	T	C	T	P161	n/a	n/a	n/a	n/a
62	7523769	3294566F6	SNP00053413	103	69	C	C	G	A8	n/d	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
62	7523769	3294839H1	SNP00099838	197	90	G	G	A	R15	0.72	0.81	n/d	0.74
62	7523769	6958246H1	SNP00053414	226	391	T	C	T	N115	n/a	n/a	n/a	n/a
62	7523769	6958246H1	SNP00073871	85	250	A	A	G	V68	0.77	n/a	n/a	n/a
62	7523769	7633850H1	SNP00132266	15	379	C	C	T	L111	n/a	n/a	n/a	n/a
62	7523769	8007681H1	SNP00040156	266	97	C	C	A	C17	0.88	0.92	n/d	0.86
62	7523769	8007681H1	SNP00132265	455	286	G	G	A	A80	n/a	n/a	n/a	n/a
63	7523785	3294566F6	SNP00053413	103	69	C	C	G	A8	n/d	n/a	n/a	n/a
63	7523785	3294839H1	SNP00099838	197	90	G	G	A	R15	0.72	0.81	n/d	0.74
63	7523785	6449492H1	SNP00040156	231	97	A	C	A	stop17	0.88	0.92	n/d	0.86
63	7523785	6449492H1	SNP00132265	420	286	A	G	A	A80	n/a	n/a	n/a	n/a
63	7523785	6449492H1	SNP00132266	513	379	C	C	T	L111	n/a	n/a	n/a	n/a
63	7523785	6958246H1	SNP00053414	226	391	T	C	T	N115	n/a	n/a	n/a	n/a
63	7523785	6958246H1	SNP00073871	85	250	A	A	G	V68	0.77	n/a	n/a	n/a
64	7523836	1262781R1	SNP00042104	434	154	C	C	T	P37	n/a	n/a	n/a	n/a
64	7523836	1262781R1	SNP00108772	337	57	C	C	T	T5	n/d	0.89	n/d	0.97
64	7523836	1262781R1	SNP00108773	529	250	G	G	A	L69	n/a	n/a	n/a	n/a
64	7523836	1299364H1	SNP00037509	61	1103	C	C	T	L354	n/a	n/a	n/a	n/a
64	7523836	1352494H1	SNP00037508	91	1072	C	C	T	T343	n/a	n/a	n/a	n/a
64	7523836	2457538F6	SNP00042104	198	153	C	C	T	P37	n/a	n/a	n/a	n/a
64	7523836	2457538F6	SNP00108773	294	249	G	G	A	R69	n/a	n/a	n/a	n/a
64	7523836	3511216F6	SNP00037508	418	1096	C	C	T	S351	n/a	n/a	n/a	n/a
64	7523836	6837889H1	SNP00124191	60	80	G	G	C	V13	n/d	n/d	n/d	n/d
64	7523836	7353813H1	SNP00108773	61	256	G	G	A	L71	n/a	n/a	n/a	n/a
64	7523836	7435949H1	SNP00108772	332	58	T	C	T	T5	n/d	0.89	n/d	0.97
64	7523836	7609996J1	SNP00042105	42	927	C	C	G	T295	n/a	n/a	n/a	n/a
64	7523836	7703523J1	SNP00042104	79	136	C	C	T	D31	n/a	n/a	n/a	n/a
64	7523836	7703523J1	SNP00108773	175	232	G	G	A	E63	n/a	n/a	n/a	n/a
64	7523836	7738579H1	SNP00037508	19	1088	C	C	T	P349	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
66	7523880	023120H1	SNP00119833	65	564	A	A	G	I175	n/d	n/d	n/d	n/d
66	7523880	062669H1	SNP00076681	23	365	C	C	T	F108	n/a	n/a	n/a	n/a
66	7523880	1416206H1	SNP00013136	6	287	G	G	A	P82	n/a	n/a	n/a	n/a
66	7523880	1519781H1	SNP00133109	68	952	T	T	C	I304	n/a	n/a	n/a	n/a
66	7523880	1688517H1	SNP00013137	150	1358	T	T	C	S439	n/a	n/a	n/a	n/a
66	7523880	7410271H1	SNP00119833	229	618	G	A	G	A193	n/d	n/d	n/d	n/d
66	7523880	7653792H1	SNP00013137	497	1357	T	T	C	I439	n/a	n/a	n/a	n/a
67	7523812	1288755T6	SNP00028951	93	2835	T	T	G	noncoding	n/a	n/a	n/a	n/a
67	7523812	1441213T6	SNP00028951	140	2793	T	T	G	noncoding	n/a	n/a	n/a	n/a
67	7523812	1482372T6	SNP00028951	130	2785	T	T	G	noncoding	n/a	n/a	n/a	n/a
67	7523812	1507357F6	SNP00108053	96	181	C	C	T	H49	n/a	n/a	n/a	n/a
67	7523812	1738590F6	SNP00028950	117	2132	G	G	A	noncoding	n/d	n/a	n/a	n/a
67	7523812	1738590T6	SNP00028951	114	2803	T	T	G	noncoding	n/a	n/a	n/a	n/a
67	7523812	1860190T6	SNP00028951	130	2784	T	T	G	noncoding	n/a	n/a	n/a	n/a
67	7523812	1876357T6	SNP00028951	122	2787	T	T	G	noncoding	n/a	n/a	n/a	n/a
67	7523812	1968762T6	SNP00028951	103	2811	T	T	G	noncoding	n/a	n/a	n/a	n/a
67	7523812	2642444F6	SNP00108053	43	180	C	C	T	D48	n/a	n/a	n/a	n/a
67	7523812	2642444T6	SNP00028951	103	2823	T	T	G	noncoding	n/a	n/a	n/a	n/a
67	7523812	2662035T6	SNP00028951	136	2798	T	T	G	noncoding	n/a	n/a	n/a	n/a
67	7523812	2666421T6	SNP00028951	120	2817	T	T	G	noncoding	n/a	n/a	n/a	n/a
67	7523812	2904758T6	SNP00028951	80	2832	T	T	G	noncoding	n/a	n/a	n/a	n/a
67	7523812	3095070T6	SNP00028951	47	2840	T	T	G	noncoding	n/a	n/a	n/a	n/a
67	7523812	390112T6	SNP00028951	74	2837	T	T	G	noncoding	n/a	n/a	n/a	n/a
69	7524357	2227450H1	SNP00144483	205	44	C	C	T	L10	n/a	n/a	n/a	n/a
69	7524357	2304848H1	SNP00121670	202	1078	T	T	G	H354	0.97	n/d	n/d	n/d
69	7524357	7659421H1	SNP00121670	406	1079	T	T	G	Y355	0.97	n/d	n/d	n/d
70	7524808	3811609H1	SNP00053165	146	1108	G	G	A	R369	n/a	n/a	n/a	n/a
70	7524808	3904940H1	SNP00053165	190	1111	G	G	A	L370	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
70	7524808	7695894J1	SNP00138060	360	657	C	C	T	T219	n/a	n/a	n/a	n/a
71	7522161	1238427R6	SNP00016370	355	1363	G	G	T	noncoding	0.97	n/a	n/a	n/a
71	7522161	1511916F6	SNP00000813	137	170	G	A	G	noncoding	0.13	0.04	n/d	0.12
71	7522161	1632855T6	SNP00016370	145	1375	T	G	T	noncoding	0.97	n/a	n/a	n/a
71	7522161	1797396R6	SNP00016369	464	472	T	T	C	S99	n/a	n/a	n/a	n/a
71	7522161	1797396T6	SNP00016370	149	1366	G	G	T	noncoding	0.97	n/a	n/a	n/a
72	7523999	3082674H1	SNP00031233	160	3833	C	C	T	T1275	n/d	n/a	n/a	n/a
73	7524024	3855123H1	SNP00049299	237	2051	T	C	T	G665	0.03	0.03	n/a	0.17
75	7524494	1693896H1	SNP00051085	133	4042	G	G	A	G1295	n/a	n/a	n/a	n/a
75	7524494	6824360H1	SNP00072192	300	1727	G	G	A	R523	0.48	n/a	n/a	n/a
76	7524965	1796202H1	SNP00074911	132	1220	T	T	G	noncoding	n/d	n/d	n/d	n/d
76	7524965	1796202H1	SNP00074912	139	1227	T	T	G	noncoding	n/d	n/d	n/d	n/d
76	7524965	2514928F6	SNP00074910	147	815	C	C	T	noncoding	n/d	n/d	n/d	n/d
76	7524965	4266042H1	SNP00039704	62	483	C	C	T	noncoding	n/a	n/a	n/a	n/a
76	7524965	5974372H1	SNP00148014	155	564	C	C	T	noncoding	n/a	n/a	n/a	n/a
76	7524965	7131217H1	SNP00040492	490	1553	G	G	C	noncoding	n/d	n/a	n/a	n/a
78	7516620	1602410T6	SNP00000583	108	6759	A	A	G	noncoding	0.84	0.57	0.85	0.87
78	7516620	1620639H1	SNP00066254	182	5083	C	C	A	D1636	n/d	n/d	n/d	n/d
78	7516620	1694123H1	SNP00000583	115	6658	A	A	G	noncoding	0.84	0.57	0.85	0.87
78	7516620	1781418H1	SNP00015897	260	6829	T	T	C	noncoding	n/d	n/d	n/d	n/d
78	7516620	3451862R6	SNP00066254	331	5082	C	C	A	A1636	n/d	n/d	n/d	n/d
78	7516620	3451862R6	SNP00104328	386	5137	C	C	T	Y1654	n/d	n/a	n/a	n/a
78	7516620	3451862T6	SNP00000583	162	6706	A	A	G	noncoding	0.84	0.57	0.85	0.87
78	7516620	3489438H1	SNP00104328	106	5138	C	C	T	L1655	n/d	n/a	n/a	n/a
78	7516620	4032535H1	SNP00104327	228	3637	G	G	T	A1154	0.67	n/a	n/a	n/a
78	7516620	6616687H1	SNP00136772	415	2764	C	C	G	D863	n/a	n/a	n/a	n/a
78	7516620	6937224H1	SNP00066253	55	2464	T	C	T	F763	n/a	n/a	n/a	n/a
78	7516620	7227961H1	SNP00121500	399	5323	G	G	C	K1716	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
78	7516620	7227961H1	SNP00136773	392	5316	T	C	T	I1714	n/a	n/a	n/a	n/a
78	7516620	7595083H1	SNP00136772	190	2763	C	C	G	A863	n/a	n/a	n/a	n/a
78	7516620	7744392H1	SNP00000583	316	6659	A	A	G	noncoding	0.84	0.57	0.85	0.87
78	7516620	7744392H1	SNP00015897	143	6830	T	T	C	noncoding	n/d	n/d	n/d	n/d
78	7516620	7744392J1	SNP00066254	195	5084	C	C	A	H1637	n/d	n/d	n/d	n/d
78	7516620	7744392J1	SNP00104328	250	5139	T	C	T	L1655	n/d	n/a	n/a	n/a
79	7525149	7047595H1	SNP00140763	430	133	G	G	A	E34	n/a	n/a	n/a	n/a
79	7525149	7453323H1	SNP00099442	537	384	A	G	A	V117	n/a	n/a	n/a	n/a
80	7513047	1241577H1	SNP00050432	39	254	C	C	T	A58	n/a	n/a	n/a	n/a
80	7513047	1450595H1	SNP00066499	23	233	C	C	T	P51	n/a	n/a	n/a	n/a
80	7513047	1605052H1	SNP00047370	101	168	T	T	C	L29	n/a	n/a	n/a	n/a
80	7513047	2353756T6	SNP00050432	376	253	C	C	T	P58	n/a	n/a	n/a	n/a
81	7513056	1570466H1	SNP00052052	180	702	T	T	C	noncoding	n/a	n/a	n/a	n/a
81	7513056	1694538H1	SNP00052051	104	492	T	T	G	noncoding	n/d	n/a	n/a	n/a
81	7513056	1996620R6	SNP00052052	132	703	T	T	C	noncoding	n/a	n/a	n/a	n/a
81	7513056	2205425T6	SNP00052052	294	779	T	T	C	noncoding	n/a	n/a	n/a	n/a
82	7513245	2170887H1	SNP00014843	84	2254	G	G	A	noncoding	n/a	n/a	n/a	n/a
82	7513245	2715147H1	SNP00129452	150	2031	G	G	C	noncoding	n/a	n/a	n/a	n/a
82	7513245	4322825H1	SNP00014843	120	2242	A	G	A	noncoding	n/a	n/a	n/a	n/a
83	7513711	1453384H1	SNP00021324	83	623	G	G	A	noncoding	n/a	n/a	n/a	n/a
83	7513711	1834689H1	SNP00021325	75	856	C	C	A	noncoding	n/a	n/a	n/a	n/a
83	7513711	5751740H1	SNP00116782	419	1310	A	G	A	noncoding	n/a	n/a	n/a	n/a
85	7513969	1286518H1	SNP00116511	23	170	T	T	C	V55	n/a	n/a	n/a	n/a
85	7513969	1383575F6	SNP00116513	181	872	G	G	C	S289	n/a	n/a	n/a	n/a
85	7513969	1511773F6	SNP00116512	270	576	A	A	G	K190	n/a	n/a	n/a	n/a
85	7513969	2888824H1	SNP00068575	81	357	C	C	T	A117	n/a	n/a	n/a	n/a
85	7513969	5402433H1	SNP00116513	129	864	G	G	C	Q286	n/a	n/a	n/a	n/a
85	7513969	7077356H1	SNP00068575	315	376	T	C	T	L124	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
85	7513969	758129H1	SNP00049630	10	1169	A	A	G	Q388	n/a	n/a	n/a	n/a
85	7513969	7631079J1	SNP00068575	398	359	C	C	T	A118	n/a	n/a	n/a	n/a
86	7512119	1306758F6	SNP00098047	175	220	T	T	C	I22	n/d	n/d	n/d	n/d
86	7512119	1306758H1	SNP00098047	175	219	T	T	C	I22	n/d	n/d	n/d	n/d
86	7512119	1781932H1	SNP00051598	230	1583	G	G	A	noncoding	n/a	n/a	n/a	n/a
86	7512119	2493973H1	SNP00098048	231	277	A	G	A	Q41	n/a	n/a	n/a	n/a
86	7512119	2731281H1	SNP00098049	61	998	T	T	C	L282	0.53	0.21	0.7	0.35
86	7512119	8533159H1	SNP00098049	305	1002	C	T	C	P283	0.53	0.21	0.7	0.35
87	7515577	1913012H1	SNP00144527	16	78	C	C	G	P13	n/a	n/a	n/a	n/a
87	7515577	8610814H1	SNP00144528	195	260	G	G	A	L73	n/a	n/a	n/a	n/a
87	7515577	8610814J1	SNP00019600	20	790	C	C	T	noncoding	n/a	n/a	n/a	n/a
89	7513838	1288755T6	SNP00028951	93	3434	T	T	G	noncoding	n/a	n/a	n/a	n/a
89	7513838	1441213T6	SNP00028951	140	3392	T	T	G	noncoding	n/a	n/a	n/a	n/a
89	7513838	1482372T6	SNP00028951	130	3384	T	T	G	noncoding	n/a	n/a	n/a	n/a
89	7513838	1738590F6	SNP00028950	117	2731	G	G	A	noncoding	n/d	n/a	n/a	n/a
89	7513838	1738590T6	SNP00028951	114	3402	T	T	G	noncoding	n/a	n/a	n/a	n/a
89	7513838	1860190T6	SNP00028951	130	3383	T	T	G	noncoding	n/a	n/a	n/a	n/a
89	7513838	1876357T6	SNP00028951	122	3386	T	T	G	noncoding	n/a	n/a	n/a	n/a
89	7513838	1968762T6	SNP00028951	103	3410	T	T	G	noncoding	n/a	n/a	n/a	n/a
89	7513838	2642444F6	SNP00108053	43	347	C	C	T	D48	n/a	n/a	n/a	n/a
89	7513838	2642444T6	SNP00028951	103	3422	T	T	G	noncoding	n/a	n/a	n/a	n/a
89	7513838	2662035T6	SNP00028951	136	3397	T	T	G	noncoding	n/a	n/a	n/a	n/a
89	7513838	2666421T6	SNP00028951	120	3416	T	T	G	noncoding	n/a	n/a	n/a	n/a
89	7513838	2904758T6	SNP00028951	80	3431	T	T	G	noncoding	n/a	n/a	n/a	n/a
89	7513838	2994263T6	SNP00028951	48	3468	T	T	G	noncoding	n/a	n/a	n/a	n/a
89	7513838	3095070T6	SNP00028951	47	3439	T	T	G	noncoding	n/a	n/a	n/a	n/a
89	7513838	390112T6	SNP00028951	74	3436	T	T	G	noncoding	n/a	n/a	n/a	n/a
90	7515163	488790H1	SNP00073287	125	1441	C	C	T	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
90	7515163	488790R7	SNP00073287	125	1445	C	C	T	noncoding	n/a	n/a	n/a	n/a
91	7516929	2665352F6	SNP00029431	62	2700	G	G	C	K729	n/d	n/d	n/d	n/d
91	7516929	2859393F6	SNP00137029	155	2154	G	G	A	L547	0.98	n/a	n/a	n/a
91	7516929	2859454H1	SNP00006205	126	1791	T	C	T	D426	0.62	n/a	n/a	n/a
91	7516929	3201818H1	SNP00138993	31	2123	A	A	G	D537	n/a	n/a	n/a	n/a
91	7516929	6507178H1	SNP00022929	561	472	C	C	G	noncoding	n/a	n/a	n/a	n/a
91	7516929	6771008H1	SNP00115918	465	2526	C	C	T	S671	n/d	n/a	n/a	n/a
91	7516929	7640070H2	SNP00029431	167	2697	G	G	C	L728	n/d	n/d	n/d	n/d
92	7515570	1492587H1	SNP00066276	31	355	C	C	T	Q114	n/a	n/a	n/a	n/a
92	7515570	7742874H1	SNP00066276	277	343	C	C	T	R110	n/a	n/a	n/a	n/a
92	7515570	8616944H1	SNP00062208	75	4	G	G	A	noncoding	n/a	n/a	n/a	n/a
93	7515680	066709H1	SNP00016107	127	305	C	C	A	S101	n/a	n/a	n/a	n/a
93	7515680	066709H1	SNP00016108	158	336	A	A	G	G111	n/a	n/a	n/a	n/a
93	7515680	7627666J1	SNP00016107	57	307	C	C	A	Q102	n/a	n/a	n/a	n/a
93	7515680	7627666J1	SNP00016108	88	338	A	A	G	K112	n/a	n/a	n/a	n/a
95	7517501	1449338F6	SNP00017078	229	3824	C	C	A	noncoding	n/a	n/a	n/a	n/a
95	7517501	1449338T6	SNP00017078	2	3817	C	C	A	noncoding	n/a	n/a	n/a	n/a
95	7517501	2312858H1	SNP00001264	221	3604	C	C	T	noncoding	n/a	n/a	n/a	n/a
95	7517501	2312858H1	SNP00127766	114	3497	G	G	C	noncoding	n/a	n/a	n/a	n/a
95	7517501	2641285H1	SNP00017076	136	156	C	C	T	P28	n/a	n/a	n/a	n/a
95	7517501	6389009H1	SNP00017076	191	159	C	C	T	D29	n/a	n/a	n/a	n/a
97	7518626	1811146F6	SNP00040899	407	636	G	G	A	G212	n/d	n/d	n/d	n/d
97	7518626	1811146F6	SNP00107474	296	525	G	G	A	G175	0.73	n/a	n/a	n/a
97	7518626	2043808H1	SNP00107475	76	703	G	C	G	T234	n/d	n/a	n/a	n/a
97	7518626	2124195H1	SNP00040899	127	635	G	G	A	E212	n/d	n/d	n/d	n/d
97	7518626	2124195H1	SNP00107474	16	524	G	G	A	G175	0.73	n/a	n/a	n/a
97	7518626	2500280T6	SNP00040899	323	666	G	G	A	G222	n/d	n/d	n/d	n/d
97	7518626	2500280T6	SNP00107474	434	555	G	G	A	G185	0.73	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
97	7518626	2665153T6	SNP00040899	335	672	G	G	A	R224	n/d	n/d	n/d	n/d
97	7518626	2721711T6	SNP00040899	351	643	G	G	A	L214	n/d	n/d	n/d	n/d
97	7518626	2721711T6	SNP00107474	462	532	G	G	A	Q177	0.73	n/a	n/a	n/a
97	7518626	2721711T6	SNP00107475	283	711	C	C	G	P237	n/d	n/a	n/a	n/a
97	7518626	4548869T1	SNP00040899	353	638	G	G	A	G213	n/d	n/d	n/d	n/d
97	7518626	4548869T1	SNP00107474	465	526	A	G	A	G175	0.73	n/a	n/a	n/a
97	7518626	4548869T1	SNP00107475	285	706	G	C	G	S235	n/d	n/a	n/a	n/a
97	7518626	7752241H1	SNP00107475	351	704	C	C	G	P235	n/d	n/a	n/a	n/a
98	7515714	1466516F6	SNP00050171	279	537	A	A	G	noncoding	0.26	n/a	n/a	n/a
98	7515714	2462908H1	SNP00050169	174	247	G	G	A	R74	0.5	n/a	n/a	n/a